Successive Refinement with Decoder Cooperation and its Channel Coding Duals

Himanshu Asnani†, Haim Permuter* and Tsachy Weissman†

Abstract

We study cooperation in multi terminal source coding models involving successive refinement. Specifically, we study the case of a single encoder and two decoders, where the encoder provides a common description to both the decoders and a private description to only one of the decoders. The decoders cooperate via *cribbing*, i.e., the decoder with access only to the common description is allowed to observe, in addition, a deterministic function of the reconstruction symbols produced by the other. We characterize the fundamental performance limits in the respective settings of non-causal, strictly-causal and causal cribbing. We use a new coding scheme, referred to as *Forward Encoding and Block Markov Decoding*, which is a variant of one recently used by Cuff and Zhao for coordination via implicit communication. Finally, we use the insight gained to introduce and solve some dual channel coding scenarios involving Multiple Access Channels with cribbing.

Index Terms

Block Markov Decoding, Conferencing, Cooperation, Coordination, Cribbing, Double Binning, Duality, Forward Encoding, Joint Typicality, Successive Refinement.

I. Introduction

Cooperation can dramatically boost the performance of a network. The literature abounds with models for cooperation, when communication between nodes of a network is over a noisy channel. In multiple access channels, the setting of *cribbing* was introduced by Willems and Van der Muelen in [1], where one encoder obtains the channel input symbols of the other encoder (referred to as "crib") and uses it

†Department of Electrical Engineering, Stanford University

*Department of Electrical Engineering, Ben Gurion University

for coding over a multiple access channel (MAC). This was further generalized to deterministic function cribbing (where an encoder obtains a deterministic function of the channel input symbols of another encoder) and to cribbing with actions (where one encoder can control the quality and availability of the "crib" by taking cost constrained actions) by Permuter and Asnani in [2]. Cooperation can also be modeled as information exchange among the transmitters and receivers via rate limited links, generally referred to as *conferencing* in the literature. Such a model was introduced in the context of the MAC by Willems in [3], and subsequently studied by by Bross, Lapidoth and Wigger [4], Wiese et al. [5], Simeone et al. [6], and Maric, Yates and Kramer [7]. Cooperation has also been modeled via *conferencing/cribbing* in cognitive interference channels, such as the settings in Bross, Steinberg and Tinguely [8] and Prabhakaran and Vishwanath [9]-[10]. We refer to Ng and Goldsmith [11] for a survey of various cooperation strategies and their fundamental limits in wireless networks.

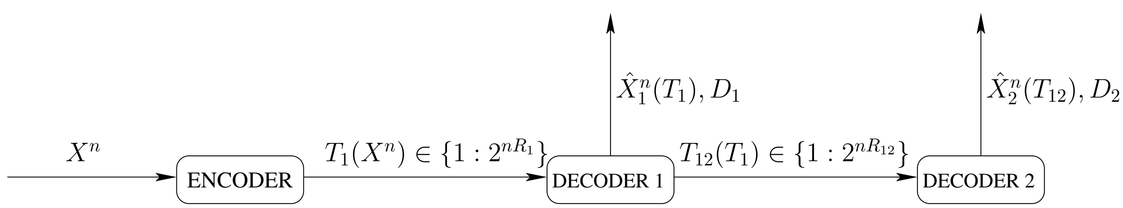

Fig. 1. Cascade source coding setup.

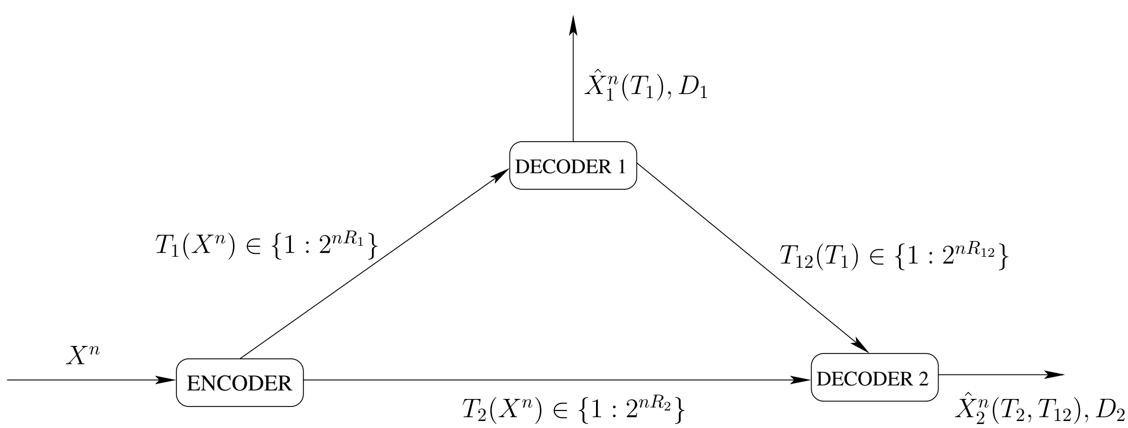

Fig. 2. Triangular source coding setup.

In multi terminal source coding, cooperation is generally modeled as a rate limited link such as in the cascade source coding setting of Yamamoto [12], Cuff, Su and El Gamal [13], Permuter and Weissman [14], Chia, Permuter and Weissman [15], as well as the triangular source coding problems of Yamamoto [16], Chia, Permuter and Weissman [15]. In cascade source coding (Fig. 1), Decoder 1 sends a description (T_{12}) to Decoder 2, which does not receive any direct description from the encoder, while in triangular source coding (Fig. 2), Decoder 1 provides a description (T_{12}) to Decoder 2 in addition to the direct description (T_2) from the encoder.

The contribution of this paper is to introduce new models of cooperation in multi terminal source coding, inspired by the *cribbing* of Willems and Van der Muelen [1] and by the implicit communication model of Cuff and Zhao [17]. Specifically, we consider cooperation between decoders in a successive refinement setting (introduced in Equitz and Cover [18]). In successive refinement, a single encoder describes a common rate to both the decoders and a private rate to only one of the decoders. We generalize this model to accommodate *cooperation* among the decoders as follows :

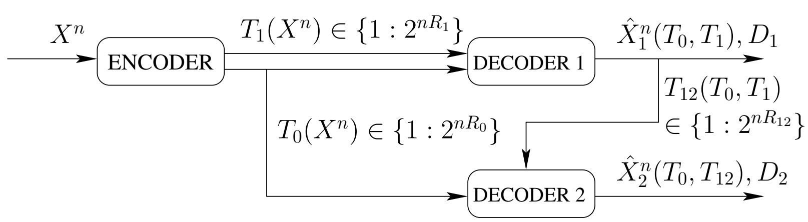

Fig. 3. Successive refinement, with decoders *cooperating* via *conferencing*.

1) *Cooperation via Conferencing* : One such cooperation model considered is that shown in Fig. 3, where the encoder provides a common description (T_0) to both the decoders and a refined description (T_1) to Decoder 1, Decoder 1 cooperates with Decoder 2 by providing an additional description (T_{12}) which is the function of its own private description (T_1), as well as the common description (T_0). This setting is inspired by the *conferencing* problem in channel coding described earlier. The region of achievable rates and distortions for this problem is given by,

$$R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2) \tag{1}$$

$$R_0 + R_{12} \geq I(X; \hat{X}_2), \tag{2}$$

for some joint probability distribution $P_{X,\hat{X}_1,\hat{X}_2}$ such that $\mathsf{E}[d_i(X_i,\hat{X}_i)] \le D_i$, for $i = 1,2$, where d_i refers to the distortion function and D_i are the distortion constraints, as is formally explained in Section II. The direct part of this characterization, namely that this region is achievable, follows standard arguments that generalize those used in the original successive refinement problem [18] (cf. Appendix A).

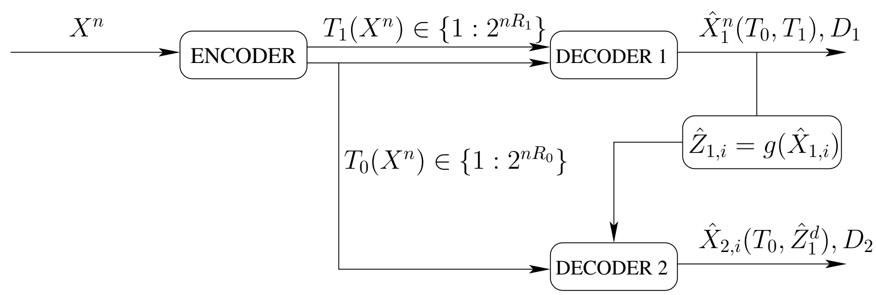

Fig. 4. Successive refinement, with decoders *cooperating* via *cribbing*. $d = n$, $d = i-1$ and $d = i$ respectively correspond to non-causal, strictly-causal and causal cribbing.

2) *Cooperation via Cribbing* : The main setting analyzed in this paper is shown in Fig. 4. A single encoder describes a common message T_0 to both decoders and a refined message T_1 to only Decoder 1. Instead of cooperating via a rate limited link, as in Fig. 3, Decoder 2 "cribs" (in the spirit of Willems and Van der Muelen [1]) a deterministic function g of the reconstruction symbols of Decoder 1, non-causally, strictly-causally, or causally. Note a trivial g function corresponds to the original successive refinement setting characterized in Equitz and Cover [18]. The goal is to find the optimal encoding and decoding strategy and to characterize the optimal encoding rate region which is defined as the set of achievable rate tuples (R_0, R_1) such that the distortion constraints are satisfied at both the decoders. Cuff and Zhao [17], considered the problem of characterizing the coordination region (non-causal, strictly causal and causal coordination) in our setting of Fig. 4, for a specific function, g, such that $g(\hat{X}_1) = \hat{X}_1$ and for a specific rate tuple $(R_0, R_1) = (0, \infty)$, that is Decoder 1 has access to the source sequence X^n while Decoder 2 uses the reconstruction symbols of Decoder 1 (non-causally, strictly-causally or causally) to estimate the source. We use a new source coding scheme which we refer to as *Forward Encoding and Block Markov Decoding*, and show that it achieves the optimal rate region for strictly causal and causal cribbing. It draws on the achievability ideas (for

causal coordination) introduced in Cuff and Zhao [17]. This scheme operates in blocks, where in the current block, the encoder encodes for the source sequence of the future block, (hence the name *Forward Encoding*) and the decoders rely on the decodability in the previous block to decode in the current block (hence the name *Block Markov Decoding*). More details about this scheme are deferred to Section III.

The general motivation for our work is an attempt to understand fundamental limits in source coding scenarios involving the availability of side information in the form of a lossily compressed version of the source. This is a departure from the standard and well studied models where side information is merely a correlated “noisy version” of the source, and is challenging because the effective ‘channel’ from source to side information is now induced by a compression scheme. Thus, rather than dictated by nature, the side information is now another degree of freedom in the design. There is no shortage of practical scenarios that motivate our models.

One such scenario may arise in the context of video coding, as considered by Aaron, Varodayan and Girod in ([19]). Consider two consecutive frames in a video file, denoted by Frame 1 and Frame 2, respectively. The video encoder starts by encoding Frame 1, and then it encodes the difference between Frame 1 and Frame 2. Decoder 1 represents decoding of Frame 1, while Decoder 2 uses the knowledge of decoded Frame 1 (via cribbing) to estimate the next frame, Frame 2.

Our problem setting is equally natural for capturing noncooperation as it is for capturing cooperation, by requiring the relevant distortions to be bounded from below rather than above (which, in turn, can be converted to our standard form of an upper bound on the distortion by changing the sign of the distortion criterion). For instance, Decoder 1 can represent an end-user with refined information (common and private rate) about a secret document, the source in our problem, while Decoder 2 has a crude information about the document (via the common rate). Decoder 1 is required to publicly announce an approximate version of the document, but due to privacy issues would like to remain somewhat cryptic about the source (as measured in terms of distortion with respect to the source) while also helping (via conferencing or cribbing) Decoder 2 to better estimate the source. For example, Decoder 1 can represent a Government agency required by law to publicly reveal features of the data, while on the other hand there are agents who make use of this publicly announced information, along with crude information about the source that they too, not only the

government, are allowed to access, to decipher or get a good estimate of the classified information (the source).

The contribution of this paper is two-fold. First, we introduce new models of decoder cooperation in source coding problems such as successive refinement, where decoders cooperate via cribbing, and we characterize the fundamental limits on performance for these problems using new classes of schemes for the achievability part. Second, we leverage the insights gained from these problems to introduce and solve a new class of channel coding scenarios that are dual to the source coding ones. Specifically, we consider the MAC with cribbing and a common message, where there are two encoders who want to communicate messages over the MAC, one has access to its own private message, there is a common message between the two encoders, and the encoders cooperate via cribbing (non-causally, strictly causally or causally).

The paper is organized as follows. Section II gives a formal description of the problem and the main results. Section III presents achievability and converses, with non-causal, causal and strictly-causal cribbing. Some special cases of our setting and numerical examples, are studied in Section IV. Channel coding duals are considered in Section V. Finally, the paper is concluded in Section VI.

II. Problem Definitions and Main Results

We begin by explaining the notation to be used throughout this paper. Let upper case, lower case, and calligraphic letters denote, respectively, random variables, specific or deterministic values which random variables may assume, and their alphabets. For two jointly distributed random variables, X and Y, let P_X, P_{XY} and $P_{X|Y}$ respectively denote the marginal of X, joint distribution of (X,Y) and conditional distribution of X given Y. X_m^n is a shorthand for the $n-m+1$ tuple $\{X_m, X_{m+1}, \cdots, X_{n-1}, X_n\}$. We impose the assumption of finiteness of cardinality on all alphabets, unless otherwise indicated.

In this section we formally define the problem considered in this paper (cf. Fig. 4). The source sequence $X_i \in \mathcal{X}, i = 1,2,...$ is drawn i.i.d. $\sim P_X$. Let $\hat{\mathcal{X}}_1$ and $\hat{\mathcal{X}}_2$ denote the reconstruction alphabets, and $d_i : \mathcal{X} \times \hat{\mathcal{X}}_i \to [0,\infty)$, for $i = 1,2$ denote single letter distortion measures. Distortion between sequences is defined in the usual way,

$$d_i(x^n, \hat{x}_i^n) = \frac{1}{n}\sum_{j=1}^{n} d_i(x_j, \hat{x}_{i,j}), \text{ for } i = 1,2. \tag{3}$$

Definition 1. *A ($2^{nR_0}, 2^{nR_1}, n$) rate-distortion code consists of the following,*

1) *Encoder,* $f_{0,n} : \mathcal{X}^n \to \{1, ..., 2^{nR_0}\}$, $f_{1,n} : \mathcal{X}^n \to \{1, ..., 2^{nR_1}\}$.
2) *Decoder 1,* $g_{1,n} : \{1, ..., 2^{nR_0}\} \times \{1, ..., 2^{nR_1}\} \to \hat{\mathcal{X}}_1^n$.
3) *Decoder 2 (depending on d in Fig. 4, the decoder mapping changes as below),*

$$g_{2,i}^{nc} : \{1, ..., 2^{nR_0}\} \times \hat{\mathcal{X}}_1^n \to \hat{\mathcal{X}}_2 \quad \textit{non-causal cribbing, } d = n \tag{4}$$

$$g_{2,i}^{sc} : \{1, ..., 2^{nR_0}\} \times \hat{\mathcal{X}}_1^{i-1} \to \hat{\mathcal{X}}_2 \quad \textit{strictly-causal cribbing, } d = i-1, \tag{5}$$

$$g_{2,i}^{c} : \{1, ..., 2^{nR_0}\} \times \hat{\mathcal{X}}_1^i \to \hat{\mathcal{X}}_2 \quad \textit{causal cribbing, } d = i \tag{6}$$

$\forall\ i = 1, ..., n$.

Definition 2. *A rate-distortion tuple (R_0, R_1, D_1, D_2) is said to be achievable if $\forall\ \epsilon > 0$, $\exists\ n$ and ($2^{nR_0}, 2^{nR_1}, n$) rate-distortion code such that the expected distortion for decoders are bounded as,*

$$\mathsf{E}\left[d_i(X_i^n, \hat{X}_i^n)\right] \leq D_i + \epsilon,\ i = 1, 2. \tag{7}$$

Definition 3. *The rate-distortion region $\mathcal{R}(D_1, D_2)$ is defined as the closure of the set to all achievable rate-distortion tuples (R_0, R_1, D_1, D_2).*

Our main results for this setting are presented in the Table I. Note that in all the rate regions in the table, we use the notation $\{a\}^+$ for $\max(a, 0)$, and we omit the distortion condition $\mathsf{E}[d_i(X_i, \hat{X}_i] \leq D_i$, $i = 1, 2$ for the sake of brevity. These results will be derived later in Section III. As another contribution, in Section V, we establish duality between the problem of successive refinement with cribbing decoders and communication over multiple access channels with cribbing encoders and a common message. We establish a complete duality between the settings (in a sense that is detailed in Section V) and rate regions of one can be obtained from those of the other by listed transformations.

Lemma 1 (Equivalence to Cascade Source Coding with Cribbing Decoders)**.** *The setup in Fig. 4 is equivalent to a cascade source coding setup with cribbing decoders as in Fig. 5 in the following way : fix a distortion pair (D_1, D_2) and let $\mathcal{R}(D_1, D_2)$ denote the rate region for the problem of successive refinement with cribbing with achievable rate pairs (R_0, R_1). Let $\tilde{R}(D_1, D_2)$ denote the closure of rate pairs, $(R_0, R_0 + R_1)$ and $\mathcal{R}_{cascade}(D_1, D_2)$ denote the rate region for the problem of cascade*

$\mathcal{R}(D_1, D_2)$	Perfect Cribbing $g(\hat{X}_1) = \hat{X}_1$	Deterministic Function Cribbing
Non-Causal $(d = n)$	(Theorem 1) $R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2)$ $R_0 \geq \{I(X; \hat{X}_1, \hat{X}_2) - H(\hat{X}_1)\}^+$ *(p.m.f.)* : $P(X, \hat{X}_1, \hat{X}_2)$	(Theorem 2) $R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2)$ $R_0 \geq \{I(X; \hat{Z}_1, \hat{X}_2) - H(\hat{Z}_1)\}^+$ *(p.m.f.)* : $P(X, \hat{X}_1, \hat{X}_2)\mathbf{1}_{\{\hat{Z}_1 = g(\hat{X}_1)\}}$
Strictly-Causal $(d = i - 1)$	(Theorem 3) $R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2)$ $R_0 \geq \{I(X; \hat{X}_1, \hat{X}_2) - H(\hat{X}_1\|\hat{X}_2)\}^+$ *(p.m.f.)* : $P(X, \hat{X}_1, \hat{X}_2)$	(Theorem 4) $R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2)$ $R_0 \geq \{I(X; \hat{Z}_1, \hat{X}_2) - H(\hat{Z}_1\|\hat{X}_2)\}^+$ *(p.m.f.)* : $P(X, \hat{X}_1, \hat{X}_2)\mathbf{1}_{\{\hat{Z}_1 = g(\hat{X}_1)\}}$
Causal $(d = i)$	(Theorem 5) $R_0 + R_1 \geq I(X; \hat{X}_1, U)$ $R_0 \geq \{I(X; \hat{X}_1, U) - H(\hat{X}_1\|U)\}^+$ *(p.m.f.)* : $P(X, \hat{X}_1, U)\mathbf{1}_{\{\hat{X}_2 = f(U)\}}$ $\|\mathcal{U}\| \leq \|\mathcal{X}\|\,\|\mathcal{X}_1\| + 4$	(Theorem 6) $R_0 + R_1 \geq I(X; \hat{X}_1, U)$ $R_0 \geq \{I(X; \hat{Z}_1, U) - H(\hat{Z}_1\|U)\}^+$ *(p.m.f.)* : $P(X, \hat{X}_1, U)\mathbf{1}_{\{\hat{Z}_1 = g(\hat{X}_1), \hat{X}_2 = f(\hat{X}_1, U)\}}$ $\|\mathcal{U}\| \leq \|\mathcal{X}\|\,\|\mathcal{X}_1\| + 4$

TABLE I
MAIN RESULTS OF THE PAPER

source coding with cribbing (closure of achievable rate pairs (R_{12}, R_1)). We then have the equivalence, $\tilde{\mathcal{R}}(D_1, D_2) = \mathcal{R}_{cascade}(D_1, D_2)$.

Proof: Proof is similar to the proof of Theorem 3 in Vasudevan, Tian and Diggavi [20]. We state it in Appendix B for quick reference. ∎

We use certain standard techniques such as Typical Average Lemma, Covering Lemma and Packing Lemma which are stated and established in [21]. Herein, we state them for the sake of quick reference. For typical sets we use the definition as in chapter 2 of [21]. Henceforth, we omit the alphabets from the notation of typical set when it is clear from context, e.g. $\mathcal{T}_\epsilon^n(X, \hat{X}_2)$ is denoted by $\mathcal{T}_\epsilon^n$.

Lemma 2 (Typical Average Lemma, Chapter 2, [21])**.** *Let $x^n \in T_\epsilon^n$. Then for any nonnegative function*

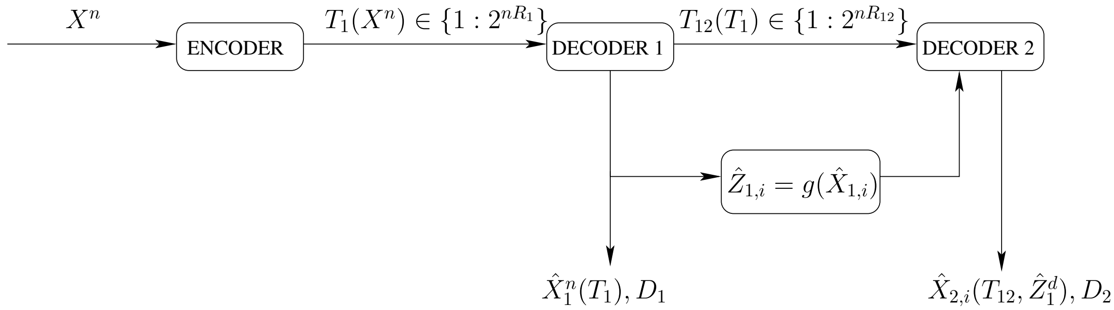

Fig. 5. Cascade source coding with *cribbing* decoders, $d = n$, $d = i-1$ and $d = i$ respectively correspond to non-causal, strictly-causal and causal cribbing.

$g(x)$ on $\mathcal{X}$,

$$(1-\epsilon)\mathsf{E}[g(X)] \leq \frac{1}{n}\sum_{i=1}^{n} g(x_i) \leq (1+\epsilon)\mathsf{E}[g(X)]. \tag{8}$$

Lemma 3 (Covering Lemma, Chapter 3, [21])**.** *Let $(U, X, \hat{X}) \sim p(u, x, \hat{x})$. Let $(U^n, X^n) \sim p(u^n, x^n)$ be a pair of arbitrarily distributed random sequences such that $P\{(U^n, X^n) \in T_\epsilon^n\} \to 1$ as $n \to \infty$ and let $\hat{X}^n(m), m \in \mathcal{A}$, where $|\mathcal{A}| \geq 2^{nR}$, be random sequences, conditionally independent of each other and of X^n given U^n, each distributed according to $\prod_{i=1}^n p_{\hat{X}|U}(\hat{x}_i|u_i)$. Then, there exists $\delta(\epsilon) \to 0$ such that $P\{(U^n, X^n, \hat{X}^n(m)) \notin T_\epsilon^n \;\forall\; m \in \mathcal{A}\} \to 0$ as $n \to \infty$, if $R > I(X; \hat{X}|U) + \delta(\epsilon)$.*

Lemma 4 (Packing Lemma, Chapter 3, [21])**.** *Let $(U, X, Y) \sim p(u, x, y)$. Let $(\tilde{U}^n, \tilde{Y}^n) \sim p(\tilde{u}^n, \tilde{y}^n)$ be a pair of arbitrarily distributed random sequences (not necessarily according to $\prod_{i=1}^n p_{U,Y}(\tilde{u}_i, \tilde{y}_i)$). Let $X^n(m), m \in \mathcal{A}$, where $|\mathcal{A}| \leq 2^{nR}$, be random sequences, each distributed according to $\prod_{i=1}^n p_{\hat{X}|U}(\hat{x}_i|u_i)$. Assume that $X^n(m), m \in \mathcal{A}$, is pairwise conditionally independent of $\tilde{Y}^n$ given $\tilde{U}^n$, but is arbitrarily dependent on other $X^n(m)$ sequences. Then, there exists $\delta(\epsilon) \to 0$ such that $P\{\tilde{U}^n, X^n, \tilde{Y}^n(m)) \in T_\epsilon^n \;\forall\; m \in \mathcal{A}\} \to 0$ as $n \to \infty$, if $R < I(X; Y|U) + \delta(\epsilon)$.*

III. Successive Refinement with Cribbing Decoders

In this section we analyze the main settings considered in this paper and derive rate regions. In the various subsections to follow we will respectively study the problem of successive refinement with non-causal, strictly causal and causal cribbing. For clarity, in each subsection, we will first study the setting of

"perfect" cribbing where $\hat{Z}_{1,i} = g(\hat{X}_{1,i}) = \hat{X}_{1,i}$ and then generalize it to cribbing with any deterministic function g.

A. Non-causal Cribbing

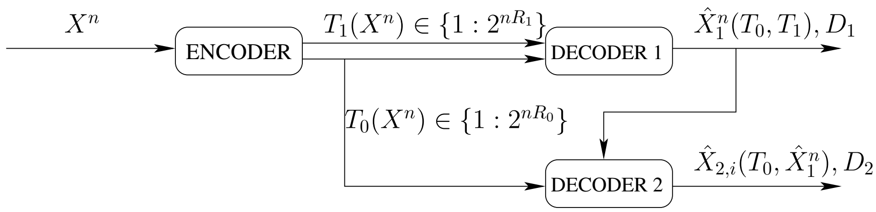

Fig. 6. Successive refinement, with decoders *cooperating* via (perfect) *non-causal cribbing*.

1) Perfect Cribbing:

Theorem 1. *The rate region $\mathcal{R}(D_1, D_2)$ for the setting in Fig. 6 with perfect (non-causal) cribbing is given as the closure of the set of all the rate tuples (R_0, R_1) such that,*

$$R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2) \tag{9}$$

$$R_0 \geq \{I(X; \hat{X}_1, \hat{X}_2) - H(\hat{X}_1)\}^+, \tag{10}$$

for some joint probability distribution $P_{X,\hat{X}_1,\hat{X}_2}$ such that $\mathsf{E}[d_i(X, \hat{X}_i)] \leq D_i$, for $i = 1, 2$.

Proof:

Achievability :

"Double Binning" scheme

Before delving into the details, we first provide a high level understanding of the achievability scheme. Consider the simplified setup where $R_0 = 0$, that is only Decoder 1 has access to the description of the source, and Decoder 2 gets the reconstruction symbols of Decoder 1 ("crib"). The intuition is to reveal a lossy description of source to the Decoder 2 through the "crib". So we first generate $2^{nI(X;\hat{X}_2)}$ $\hat{X}_2^n$ codewords, and index them as $2^{nI(X;\hat{X}_2)}$ bins. In each bin, we generate a superimposed codebook of $2^{nI(X;\hat{X}_1|\hat{X}_2)}$ $\hat{X}_1^n$ codewords. Thus total rate of $R_1 = I(X; \hat{X}_2) + I(X; \hat{X}_1|\hat{X}_2) = I(X; \hat{X}_1, \hat{X}_2)$ is needed

to describe $\hat{X}_1^n$ to Decoder 1. Decoder 2 knows $\hat{X}_1^n$ via the crib, it then needs to infer the unique bin index which was sent, as then it would infer $\hat{X}_2^n$. The only issue to verify is that the $\hat{X}_1^n$ codeword known via cribbing should not lie in two bins. We upper bound the probability of occurrence of such an event by $2^{n(I(X;\hat{X}_1,\hat{X}_2)-H(\hat{X}_1))}$, as there are overall $2^{nI(X;\hat{X}_1,\hat{X}_2)}$ $\hat{X}_1^n$ codewords, and the probability that a particular $\hat{X}_1^n$ lies in two bins is $2^{-nH(\hat{X}_1)}$. This event has a vanishing probability so long as $I(X;\hat{X}_1,\hat{X}_2) < H(\hat{X}_1)$. Thus the achieved rate region is $R_1 \geq I(X;\hat{X}_1,\hat{X}_2)$ such that the constraint $I(X;\hat{X}_1,\hat{X}_2) \leq H(\hat{X}_1)$ and distortion constraints are satisfied.

The general coding scheme when $R_0 > 0$ is depicted in Fig. 7 and has a "doubly-binned" structure. Non-zero R_0 helps reduce R_1 by providing an extra dimension of binning. We first generate $2^{nI(X;\hat{X}_2)}$ $\hat{X}_2^n$ codewords, the indexes of which are the rows (or horizontal bins), and then in each row, we generate $2^{nI(X;\hat{X}_1|\hat{X}_2)}$ $\hat{X}_1^n$ codewords. For each row, these $\hat{X}_1^n$ codewords are then binned uniformly into 2^{nR_0} vertical bins, which are the columns of our "doubly-binned" structure. Thus each bin is "doubly-indexed" (row and column index) and has a uniform number of $2^{n(I(X;\hat{X}_1|\hat{X}_2)-R_0)}$ $\hat{X}_1^n$ codewords (as in Fig. 7). Note that this extra or independent dimension of vertical binning was not there when $R_0 = 0$. Intuition is that column indexing with common rate R_0 is independent or *orthogonal* to the row indexing, and hence it helps to reduce the private rate R_1. The column or vertical bin index is described to both the decoders via common rate R_0 and thus R_1 reduces to $I(X;\hat{X}_1,\hat{X}_2) - R_0$ to describe $\hat{X}_1^n$ to Decoder 1. Here again, from knowledge of the crib, $\hat{X}_1^n$ and the column index, Decoder 2, infers the unique row index, which now will require $I(X;\hat{X}_1,\hat{X}_2) - R_0 \leq H(\hat{X}_1)$.

We now describe the achievability in full detail.

- *Codebook Generation* : Fix the distribution $P_{X,\hat{X}_1,\hat{X}_2}$, $\epsilon > 0$ such that $E[d_1(X,\hat{X}_1)] \leq \frac{D_1}{1+\epsilon}$ and $E[d_2(X,\hat{X}_2)] \leq \frac{D_2}{1+\epsilon}$. Generate codebook $\mathcal{C}_{\hat{X}_2}$ consisting of $2^{nI(X;\hat{X}_2)}$ $\hat{X}_2^n(m_h)$ codewords generated i.i.d $\sim P_{\hat{X}_2}$, $m_h \in [1 : 2^{nI(X;\hat{X}_2)}]$. For each m_h, first generate a codebook $\mathcal{C}_{\hat{X}_1}(m_h)$ consisting of $2^{nI(X;\hat{X}_1|\hat{X}_2)}$ $\hat{X}_1^n$ codewords generated i.i.d. $\sim P_{\hat{X}_1|\hat{X}_2}$, then bin them all uniformly in 2^{nR_0} vertical bins $\mathcal{B}(m_v)$, $m_v \in [1 : 2^{nR_0}]$ and in each bin index them accordingly with $l \in [1 : 2^{n(I(X;\hat{X}_1|\hat{X}_2)-R_0)}]$. As outlined earlier, m_h corresponds to the row or horizontal index and m_v corresponds to the column or vertical index in our "doubly-binned" structure, while l indexes $\hat{X}_1^n$ codewords within a "doubly-indexed" bin. Thus for each row and column index pair, (m_h, m_v), there are $2^{n(I(X;\hat{X}_1|\hat{X}_2)-R_0)}$ $\hat{X}_1^n$

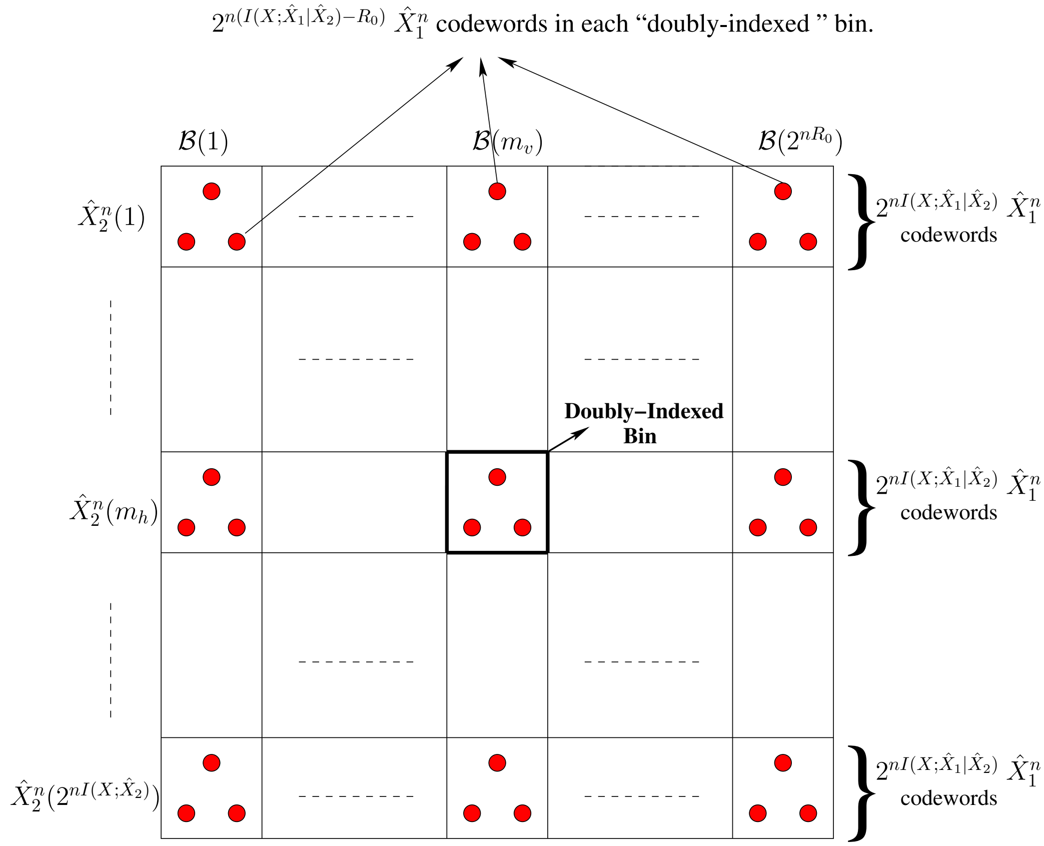

Fig. 7. "Double Binning" - achievability scheme for the non-causal perfect cribbing.

codewords. $\hat{X}_1^n$ can therefore be indexed by the triple (m_h, m_v, l). The codebooks are revealed to the encoder and both the decoders.

- *Encoding* : Given source sequence X^n, first the encoder finds m_h from $\mathcal{C}_{\hat{X}_2}$ such that $(X^n, \hat{X}_2^n(m_h)) \in \mathcal{T}_\epsilon^n$. Then the encoder finds pair (m_v, l) such that $(X^n, \hat{X}_1^n(m_h, m_v, l), \hat{X}_2^n(m_h)) \in \mathcal{T}_\epsilon^n$. Thus $\hat{X}_1^n(m_h, m_v, l) \in \mathcal{B}(m_v)$. Encoder describes column or vertical bin index m_v as R_0 to both the decoders, and the tuple (m_h, l) to the Decoder 1 as rate R_1. Thus

$$R_1 \geq I(X;\hat{X}_2) + I(X;\hat{X}_1|\hat{X}_2) - R_0 = I(X;\hat{X}_1,\hat{X}_2) - R_0. \tag{11}$$

- *Decoding* : Decoder 1 knows all the indices (m_h, m_v, l), and it constructs $\hat{X}_1^n = \hat{X}_1^n(m_h, m_v, l)$. Decoder 2 receives $\hat{X}_1^n$ from the non-causal cribbing and it also knows the column index m_v through

rate R_0. It then checks inside the column or vertical bin of index m_v, to find the unique row or horizontal bin index m_h such that $\hat{X}_1^n = \hat{X}_1^n(m_h, m_v, \tilde{l})$ for some $\tilde{l} \in [1 : 2^{n(I(X;\hat{X}_1|\hat{X}_2)-R_0)}]$. The reconstruction of the Decoder 2 is then $\hat{X}_2^n = \hat{X}_2^n(m_h)$.

- *Distortion Analysis* : Consider the following events :

1)

$$
\begin{aligned}
\mathcal{E}_1 &= \text{No } \hat{X}_2^n \text{ is jointly typical to a given } X^n && (12)\\
&= \Big\{(X^n, \hat{X}_2^n(m_h)) \notin \mathcal{T}_\epsilon^n, \forall\ m_h \in [1 : 2^{nI(X;\hat{X}_2)}]\Big\}. && (13)
\end{aligned}
$$

The probability of this event vanishes as there are $2^{nI(X;\hat{X}_2)}$ $\hat{X}_2^n$ codewords. (cf. *Covering Lemma*, Lemma 3).

2)

$$
\begin{aligned}
\mathcal{E}_2 &= \text{No } \hat{X}_1^n \text{ is jointly typical to a typical pair } (X^n, \hat{X}_2^n) && (14)\\
&= \Big\{(X^n, \hat{X}_2^n(m_h)) \in \mathcal{T}_\epsilon^n\Big\}\\
&\quad \cap\Big\{(X^n, \hat{X}_1^n(m_h, m_v, l), \hat{X}_2^n(m_h)) \notin \mathcal{T}_\epsilon^n, \forall\ m_v \in [1 : 2^{nR_0}],\\
&\quad \forall\ l \in [1 : 2^{n(I(X;\hat{X}_1|\hat{X}_2)-R_0)}]\Big\}. && (15)
\end{aligned}
$$

The probability of this event vanishes as corresponding to each m_h there are $2^{nI(X;\hat{X}_1|\hat{X}_2)}$ $\hat{X}_1^n$ codewords, (cf. *Covering Lemma*, Lemma 3). Without loss of generality, now suppose that encoder does the encoding, $(m_h, m_v, l) = (1, 1, 1)$. Decoder 2 receives $\hat{X}_1^n$ via non-causal cribbing. The next two events are with respect to Decoder 2.

3)

$$
\begin{aligned}
\mathcal{E}_3 &= \hat{X}_1^n \text{ does not lie in bin indexed by } m_h = 1 \text{ and } m_v = 1 && (16)\\
&= \Big\{\hat{X}_1^n \neq \hat{X}_1^n(1, 1, \tilde{l}), \forall\ \tilde{l} \in [1 : 2^{n(I(X;\hat{X}_1|\hat{X}_2)-R_0)}]\Big\}. && (17)
\end{aligned}
$$

But the probability of this event goes to zero, because due to our encoding procedure, $\hat{X}_1^n =$

$\hat{X}_1^n(1,1,1)$.

4)

$$\begin{aligned}\mathcal{E}_4 &= \hat{X}_1^n \text{ lies in a bin with row index, } \hat{m}_h \neq 1 \text{ and column index } m_v = 1. \quad (18)\\ &= \left\{\hat{X}_1^n = \hat{X}_1^n(\hat{m}_h, 1, \tilde{l}),\ \hat{m}_h \neq 1 \text{ for some } \tilde{l} \in [1 : 2^{n(I(X;\hat{X}_1|\hat{X}_2)-R_0)}]\right\}. \quad (19)\end{aligned}$$

Since $\hat{X}_1^n = \hat{X}_1^n(1,1,1)$, this event is equivalent to finding $\hat{X}_1^n$ lying in two different rows or horizontal bins, but with the same column or vertical bin index ($m_v = 1$). The probability of a single $\hat{X}_1^n$ codeword occurring repeatedly in two horizontal bins indexed with different row index is $2^{-nH(\hat{X}_1)}$, while knowing the column index, m_v, total number of $\hat{X}_1^n$ codewords with a particular column index are, $2^{n(I(X;\hat{X}_1,\hat{X}_2)-R_0)}$, so the probability of event $\mathcal{E}_4$ vanishes so long as,

$$I(\hat{X}_1; \hat{X}_1, \hat{X}_2) - R_0 < H(\hat{X}_1). \quad (20)$$

Thus consider the event, $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_2 \cup \mathcal{E}_4$, using the rate constraints from Eq. (11) and Eq. (30), the probability of the event vanishes if,

$$\begin{aligned}R_0 + R_1 &\geq I(X; \hat{X}_1, \hat{X}_2) \quad (21)\\ R_0 &\geq \{I(X; \hat{X}_1, \hat{X}_2) - H(\hat{X}_1)\}^+. \quad (22)\end{aligned}$$

We will now bound the distortion. Assume without loss of generality that, $d_i(\cdot,\cdot) \leq D_{max}$, for $i = 1, 2$. For both the decoders, ($i = 1, 2$),

$$\begin{aligned}E\left[d(X^n, \hat{X}_i^n)\right] &= P(\mathcal{E})E\left[d(X^n, \hat{X}_i^n)|\mathcal{E}\right] + P(\mathcal{E}^c)E\left[d(X^n, \hat{X}_i^n)|\mathcal{E}^c\right] \quad (23)\\ &\overset{(a)}{\leq} P(\mathcal{E})D_{max} + (1+\epsilon)\mathsf{E}[d(X, \hat{X}_i)] \quad (24)\\ &\leq P(\mathcal{E})D_{max} + D_i, \quad (25)\end{aligned}$$

where (a) is via typical average lemma (cf. Typical Average Lemma 2). Proof is completed by letting $n \to \infty$ when $P(\mathcal{E}) \to 0$.

Converse : Converse for this setting follows by substituting $\hat{Z}_1 = \hat{X}_1$ in the converse for the deterministic function cribbing in the next subsection.

Note 1 (Joint Typicality Decoding)**.** *Note that here our decoding for Decoder 2 relies on finding a unique bin index in which $\hat{X}_1^n$ (obtained via cribbing) lies, and there is an error if two different bins have the same $\hat{X}_1^n$. An alternative based on joint typicality decoding can also be used to achieve the same region as follows : Decoder 2 receives $\hat{X}_1^n$ via non-causal cribbing and it also knows the column index m_v through rate R_0. It then finds the unique row or horizontal bin index m_h such that $(\hat{X}_1^n, \hat{X}_1^n(m_h, m_v, \tilde{l}), \hat{X}_2^n(m_h)) \in \mathcal{T}_\epsilon^n$ for some $\tilde{l} \in [1 : 2^{n(I(X;\hat{X}_1|\hat{X}_2)-R_0)}]$. The reconstruction of the Decoder 2 is then $\hat{X}_2^n = \hat{X}_2^n(m_h)$. We analyze the following two events, assuming without loss of generality that encoder does the encoding $(m_h, m_v, l) = (1, 1, 1)$.*

- $$\begin{aligned} \mathcal{E}_{d,1} &= \text{Decoder 2 finds no jointly typical } \hat{X}_1^n \text{ indexed by } m_h = 1 \text{ and } m_v = 1 && (26) \\ &= \left\{ (\hat{X}_1^n, \hat{X}_1^n(1,1,\tilde{l}), \hat{X}_2^n(1)) \notin \mathcal{T}_\epsilon^n, \forall\, \tilde{l} \in [1 : 2^{n(I(X;\hat{X}_1|\hat{X}_2)-R_0)}] \right\}. && (27) \end{aligned}$$

  But the probability of this event goes to zero, because due to our encoding procedure, with high probability, $(X^n, \hat{X}_1^n(1,1,1), \hat{X}_2^n(1)) \in \mathcal{T}_\epsilon^n$. As $\hat{X}_1^n = \hat{X}_1^n(1,1,1)$ this implies, $(\hat{X}_1^n, \hat{X}_1^n(1,1,1), \hat{X}_2^n(1)) \in \mathcal{T}_\epsilon^n$.

- $$\begin{aligned} \mathcal{E}_{d,2} &= \text{Decoder 2 finds a jointly typical } \hat{X}_1^n \text{ codeword in row with index, } \hat{m}_h \neq 1. && (28) \\ &= \left\{ (\hat{X}_1^n, \hat{X}_1^n(\hat{m}_h, 1, \tilde{l}), \hat{X}_2^n(\hat{m}_h)) \in \mathcal{T}_\epsilon^n,\ \hat{m}_h \neq 1 \text{ for some } \tilde{l} \in [1 : 2^{n(I(X;\hat{X}_1|\hat{X}_2)-R_0)}] \right\}. && (29) \end{aligned}$$

  By Lemma 4 (Packing Lemma , substitute, $|\mathcal{A}| = 2^{n(I(X;\hat{X}_1,\hat{X}_2)-R_0)}, U = \phi, X = (\hat{X}_2, \hat{X}_1), Y = \hat{X}_1$), probability of this event goes to zero with large n, if

  $$I(\hat{X}_1; \hat{X}_1, \hat{X}_2) - R_0 \leq I(\hat{X}_1; \hat{X}_1, \hat{X}_2) = H(\hat{X}_1). \tag{30}$$

Thus we obtain the same constraint with the joint typicality decoding for Decoder 2. In all the subsections

to follow, for Decoder 2, joint typicality decoding can also be used as an alternative to the decoding that will be described.

■

2) Deterministic Function Cribbing:

Theorem 2. *The rate region $\mathcal{R}(D_1, D_2)$ for the setting in Fig. 8 with deterministic function (non-causal) cribbing is given as the closure of the set of all the rate tuples (R_0, R_1) such that,*

$$R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2) \tag{31}$$

$$R_0 \geq \{I(X; \hat{Z}_1, \hat{X}_2) - H(\hat{Z}_1)\}^+, \tag{32}$$

for some joint probability distribution $P_X P_{\hat{Z}_1, \hat{X}_2|X} P_{\hat{X}_1|\hat{Z}_1, \hat{X}_2, X}$ such that $\mathsf{E}[d_i(X, \hat{X}_i)] \leq D_i$, for $i = 1, 2$.

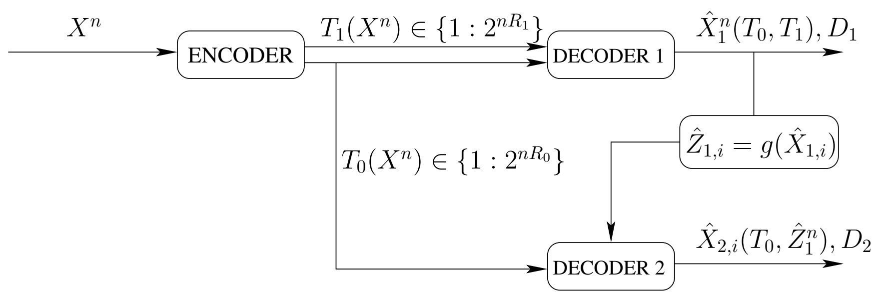

Fig. 8. Successive refinement, with decoders *cooperating* via (deterministic function) *non-causal cribbing*.

Proof:

Achievability : The scheme is similar to the achievability in the previous section, where cribbing was perfect, with some minor differences. We give an outline here and highlight the differences, deferring the complete proof to Appendix C. The codebook here also has a "doubly-binned" structure as in Fig. 7, the difference being that each "doubly-indexed" bin has a uniform number of $\hat{Z}_1^n$ codewords instead of $\hat{X}_1^n$. So first $2^{nI(X;\hat{X}_2)}$ $\hat{X}_2^n$ codewords are generated, for each of them, $2^{nI(X;\hat{Z}_1|\hat{X}_2)}$ $\hat{Z}_1^n$ codewords are generated, which are then vertically binned uniformly into 2^{nR_0} vertical bins (columns). Then for each $\hat{Z}_1^n$, $2^{nI(X;\hat{X}_1|\hat{Z}_1,\hat{X}_2)}$ $\hat{X}_1^n$ codewords are generated. Here also, the column index is described as R_0 and the remaining indices are described as R_1, which hence is equal to $I(X; \hat{X}_1, \hat{Z}_1, \hat{X}_2) - R_0 = I(X; \hat{X}_1, \hat{X}_2) -$

R_0. Decoder 1 can, as usual, construct its estimate since it knows all the indices, Decoder 2, infers the row index from the deterministic function crib, $\hat{Z}_1^n$ and knowledge of the column index. The decodability of a unique row index depends on the fact that there should not be the same $\hat{Z}_1^n$ codeword in two rows. This requires (as we saw in the previous section), $I(X; \hat{Z}_1, \hat{X}_2) - R_0 \leq H(\hat{Z}_1)$.

Converse : Assume we have a $(2^{nR_0}, 2^{nR_1}, n)$ code (as per Definition 4) achieving respective distortions D_1 and D_2. Denote $T_1 = f_{1,n}(X^n)$ and $T_0 = f_{2,n}(X^n)$. Consider,

$$
\begin{aligned}
H(\hat{Z}_1^n, T_0) &\geq I(X^n; \hat{Z}_1^n, T_0) && (33)\\
&\stackrel{(a)}{=} I(X^n; \hat{Z}_1^n, \hat{X}_2^n, T_0) && (34)\\
&\geq I(X^n; \hat{Z}_1^n, \hat{X}_2^n) && (35)\\
&= \sum_{i=1}^{n} I(X_i; \hat{Z}_1^n, \hat{X}_2^n | X^{i-1}) && (36)\\
&\stackrel{(b)}{=} \sum_{i=1}^{n} I(X_i; \hat{Z}_1^n, \hat{X}_2^n, X^{i-1}) && (37)\\
&\geq \sum_{i=1}^{n} I(X_i; \hat{Z}_{1,i}, \hat{X}_{2,i}) && (38)\\
&= n \sum_{i=1}^{n} \frac{1}{n} I(X_i; \hat{Z}_{1,i}, \hat{X}_{2,i}) && (39)\\
&\stackrel{(c)}{=} nI(X_Q; \hat{Z}_{1,Q}, \hat{X}_{2,Q} | Q) && (40)\\
&\stackrel{(d)}{=} nI(X_Q; \hat{Z}_{1,Q}, \hat{X}_{2,Q}, Q) && (41)\\
&\geq nI(X_Q; \hat{Z}_{1,Q}, \hat{X}_{2,Q}) && (42)\\
H(\hat{Z}_1^n, T_0) &\leq H(\hat{Z}_1^n) + H(T_0) && (43)\\
&\leq \sum_{i=1}^{n} H(\hat{Z}_{1,i}) + nR_0 && (44)\\
&= nH(\hat{Z}_{1,Q} | Q) + nR_0 && (45)\\
&\leq nH(\hat{Z}_{1,Q}) + nR_0 && (46)
\end{aligned}
$$

$$
\begin{aligned}
n(R_0 + R_1) &= H(T_0, T_1) && (47)\\
&= H(T_0, T_1) - H(T_0, T_1 | X^n) && (48)
\end{aligned}
$$

$$
\begin{aligned}
&= I(X^n; T_0, T_1) && (49)\\
&\overset{(e)}{=} I(X^n; T_0, T_1, \hat{X}_1^n, \hat{X}_2^n) && (50)\\
&= \sum_{i=1}^{n} I(X_i; T_0, T_1, \hat{X}_1^n, \hat{X}_2^n | X^{i-1}) && (51)\\
&\overset{(f)}{=} \sum_{i=1}^{n} I(X_i; T_0, T_1, \hat{X}_1^n, \hat{X}_2^n, X^{i-1}) && (52)\\
&\geq \sum_{i=1}^{n} I(X_i; \hat{X}_{1,i}, \hat{X}_{2,i}) && (53)\\
&= n\sum_{i=1}^{n} \frac{1}{n} I(X_i; \hat{X}_{1,i}, \hat{X}_{2,i}) && (54)\\
&= nI(X_Q; \hat{X}_{1,Q}, \hat{X}_{2,Q} | Q) && (55)\\
&= nI(X_Q; \hat{X}_{1,Q}, \hat{X}_{2,Q}, Q) && (56)\\
&\geq nI(X_Q; \hat{X}_{1,Q}, \hat{X}_{2,Q}), && (57)
\end{aligned}
$$

where (a) follows from the fact that $\hat{X}_2^n$ is a function of $(T_0, \hat{Z}_1^n)$, (b) follows from the independence of X_i and X^{i-1}, and (c) follows by defining $Q \in [1:n]$ as a uniformly distributed time sharing random variable independent of the source, (d) follows from the independence of Q with the source process, (e) follows as $(\hat{X}_1^n, \hat{X}_2^n)$ is a function of (T_0, T_1) and finally (f) follows similarly from the independence of X_i and X^{i-1}. Finally, we bound the distortion as,

$$
\begin{aligned}
D_i &\geq \mathsf{E}\left[d(X^n, \hat{X}_i^n)\right] && (58)\\
&= \mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i)\right] && (59)\\
&= \mathsf{E}[d(X_Q, \hat{X}_{i,Q})]. && (60)
\end{aligned}
$$

The proof is completed by noting that the joint distribution of $(X_Q, \hat{X}_{1,Q}, \hat{X}_{2,Q})$ is the same as that of $(X, \hat{X}_1, \hat{X}_2)$. ■

Note 2. *Due to the structure of our problem, i.e., $\hat{Z}_1 = g(\hat{X}_1)$, it is easy to prove the Markov relation, $(X, \hat{X}_2) - \hat{X}_1 - \hat{Z}_1$, hence the distribution mentioned in the statement of the theorem, can equivalently be factorized as, $P_X P_{\hat{X}_1,\hat{X}_2|X} \mathbf{1}_{\{\hat{Z}_1 = g(\hat{X}_1)\}}$, (which is the form stated in Table I). This applies similarly for theorems to follow, and we omit this explanation henceforth.*

B. Strictly-Causal Cribbing

1) Perfect Cribbing:

Theorem 3. *The rate region $\mathcal{R}(D_1, D_2)$ for the setting in Fig. 9 with perfect cribbing (strictly causal) is given by the closure of the set of all the rate tuples (R_0, R_1) such that,*

$$R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2) \tag{61}$$

$$R_0 \geq \{I(X; \hat{X}_1, \hat{X}_2) - H(\hat{X}_1|\hat{X}_2)\}^+, \tag{62}$$

for some joint probability distribution $P_{X,\hat{X}_1,\hat{X}_2}$ such that $\mathsf{E}[d_i(X, \hat{X}_i)] \leq D_i$, for $i = 1, 2$.

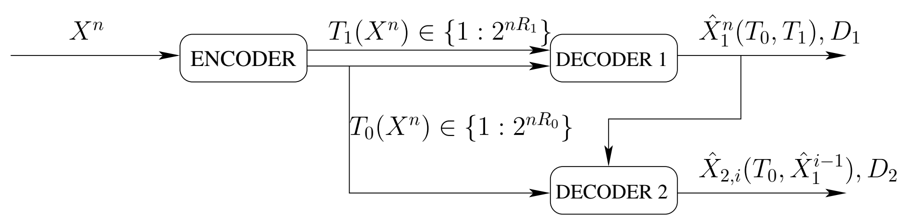

Fig. 9. Successive refinement, with decoders *cooperating* via (perfect) *strictly-causal cribbing*.

Proof:

Achievability :

We will show the achievability of the following region instead,

$$R_0 + R_1 \geq I(X; \hat{X}_1, U) \tag{63}$$

$$R_0 \geq \{I(X; \hat{X}_1, U) - H(\hat{X}_1|U)\}^+, \tag{64}$$

for some joint probability distribution $P_{X,\hat{X}_1,U}\mathbf{1}_{\{\hat{X}_2=f(U)\}}$. Note that the rate region in the theorem will then be obtained by simply taking $U = \hat{X}_2$. Here we deliberately present our encoding scheme with an auxiliary random variable as this will be used (with minor changes) to derive the achievable region for the case of causal cribbing discussed in the next subsection.

“Forward Encoding” and “Block Markov Decoding” scheme :

We use a new scheme that we refer to as “Forward Encoding” and “Block Markov Decoding”. We first briefly give an overview of the coding scheme and for simplicity consider the case when common rate $R_0 = 0$. Thus the source description is available only to Decoder 1, while Decoder 2 has access to the reconstruction symbols of Decoder 1, but only strictly-causally. Hence in principle we cannot deploy a scheme to operate in one block as was done for non-causal cribbing. We need to use a scheme to operate in multiple (large number) of blocks, and use an encoding procedure where $\hat{X}_1^n$ of the previous block carries information about the source sequence of the current block. In this way due to strictly causal cribbing, in the current block, Decoder 2 will know all the reconstruction symbols of Decoder 1 from the previous block, which will contain information about the source for the current block. This is the main idea and is operated as follows : in each block, first we generate $2^{nI(X;U)}$ U^n codewords, and for each U^n codeword, we generate $2^{nI(X;U)}$ bins and in each bin $2^{nI(X;\hat{X}_1|U)}$ $\hat{X}_1^n$ codewords are generated. In each block, U^n is jointly typical with the source sequence in the current block and the bin index describes the U^n sequence jointly typical with the source sequence of the future block. This bin index carries information about the source in the future block. Hence, we address encoding as “Forward Encoding”. Decoding is “Block Markov Decoding”, as it assumes both decoders have currently decoded the U^n sequence of the previous block. The bin index and index of the $\hat{X}_1^n$ codewords is described as R_1 which hence is taken to be $I(X;U) + I(X;\hat{X}_1|U) = I(X;\hat{X}_1, U)$. Due to cribbing, Decoder 2 knows the $\hat{X}_1^n$ of the previous block and aims to find the bin index in which it lies. And as we argued in previous sections, this is possible if $I(X;\hat{X}_1, U) \leq H(\hat{X}_1|U)$.

The general scheme when $R_0 > 0$ is depicted in Fig. 10. The additional step which we add to the description above (for $R_0 = 0$) is to bin in an extra dimension, i.e., with respect to each U^n sequence we generate a “doubly-binned” codebook (as in the achievability of non-causal cribbing, cf. Fig. 7). The row index encodes U^n sequences of the future block and $\hat{X}_1^n$ codewords for each row are uniformly binned into 2^{nR_0} columns. The column index is the common description to both decoders, so R_1 reduces to $I(X;\hat{X}_1, U) - R_0$, and the decodability of Decoder 2 requires the condition $I(X;\hat{X}_1, U) - R_0 \leq H(\hat{X}_1|U)$.

We now explain this coding scheme in detail and how it helps establish the achievable region when the cooperation between the decoders is via strictly causal cribbing.

1) *Codebook Generation :* The scheme does compression in blocks. Fix the number of blocks to be B. In each block, n source symbols are compressed. Fix a joint probability distribution, $P_{U,X,\hat{X}_1,\hat{X}_2} = P_{U,X,\hat{X}_1}\mathbf{1}_{\{\hat{X}_2=f(U)\}}$ for some function f and $\epsilon > 0$ such that $\mathsf{E}[d_1(X,\hat{X}_1)] \leq \frac{D_1}{1+\epsilon}$ and $\mathsf{E}[d_2(X,\hat{X}_2)] \leq \frac{D_2}{1+\epsilon}$.

   Now in each block we generate codebook as follows. First we generate a codebook $\mathcal{C}_U(b) = \{u^n(b,m) \sim \prod_{i=1}^{n} P_U(u_i(b,m)), m = [1:2^{nI(X;U)}]\}$ for each block $b \in [1:B]$. For each $u^n(b,m)$, we create $2^{nI(X;U)}$ horizontal bins or rows $\mathcal{B}(m_h)$ which are indexed as $m_h \in [1:2^{nI(X;U)}]$. In each bin we generate a codebook $2^{nI(X;\hat{X}_1|U)}$ $\hat{X}_1^n$ codewords which are then binned again into 2^{nR_0} vertical bins or columns, $\mathcal{B}(m_v)$ uniformly, $m_v \in [1:2^{nR_0}]$ and index them accordingly by $l \in [1:2^{n(I(X;\hat{X}_1|U)-R_0)}]$. Thus $\hat{X}_1^n$ can be equivalently indexed as the tuple (b,m,m_h,m_v,l). Hence for each u^n as explained earlier we have a "doubly-binned" structure, m_h denotes the row index and m_v denotes the column index. The codebooks are then revealed to both the encoder and decoders.

2) *Encoding :* X^{nB} is known to the encoder. From now on additional subscripts will stand for block index, eg. $m_{h,2}$ means the row index in block 2, or $m_{v,2}$ means the column index in block 2. Also additional scripts in parenthesis would denote the sequence in a block, eg. $X^n(b)$ will stand for the source sequence in block b, $\hat{X}_1^n(b)$ stands for reconstruction of Decoder 1 in block b. Encoding is as follows :

   a) For the first block, $b = 1$, assume $m_1 = 1$. Encoder then finds index m_2, such that $(X^n(2), U^n(2,m_2)) \in \mathcal{T}_\epsilon^n$. The encoder then looks in the codebook $\mathcal{C}_U(1)$ to find $U^n(1,m_1)$. Then it looks in the row or horizontal bin indexed by $m_{h,1} = m_2$ corresponding to the found $U^n(1,m_1)$, and finds the index tuple $(m_{v,1},l_1)$ such that $(\hat{X}_1^n(1,m_1,m_{h,1},m_{v,1},l_1), X^n(1), U^n(1,m_1)) \in \mathcal{T}_\epsilon^n$. As found $\hat{X}_1^n \in \mathcal{B}(m_{v,1})$, the index tuple $(m_{h,1},l_1)$ is described as R_1 and $m_{v,1}$ is described as R_0.

   b) In the block b ($\in [2:B-1]$) encoder knows m_b from encoding procedure in previous block such that $(X^n(b), U^n(b,m_b)) \in \mathcal{T}_\epsilon^n$. It then finds index m_{b+1} such that $(X^n(b+1), U^n(b+1,m_{b+1})) \in \mathcal{T}_\epsilon^n$. Now the encoder identifies the codeword, $U^n(b,m_b)$, from the codebook $\mathcal{C}_U(b)$, looks in the corresponding row or horizontal bin indexed as $m_{h,b} = m_{b+1}$ and finds the index tuple $(m_{v,b},l_b)$ such that $(\hat{X}_1^n(b,m_b,m_{h,b},m_{v,b},l_b), X^n(b), U^n(b,m_b)) \in \mathcal{T}_\epsilon^n$. As found

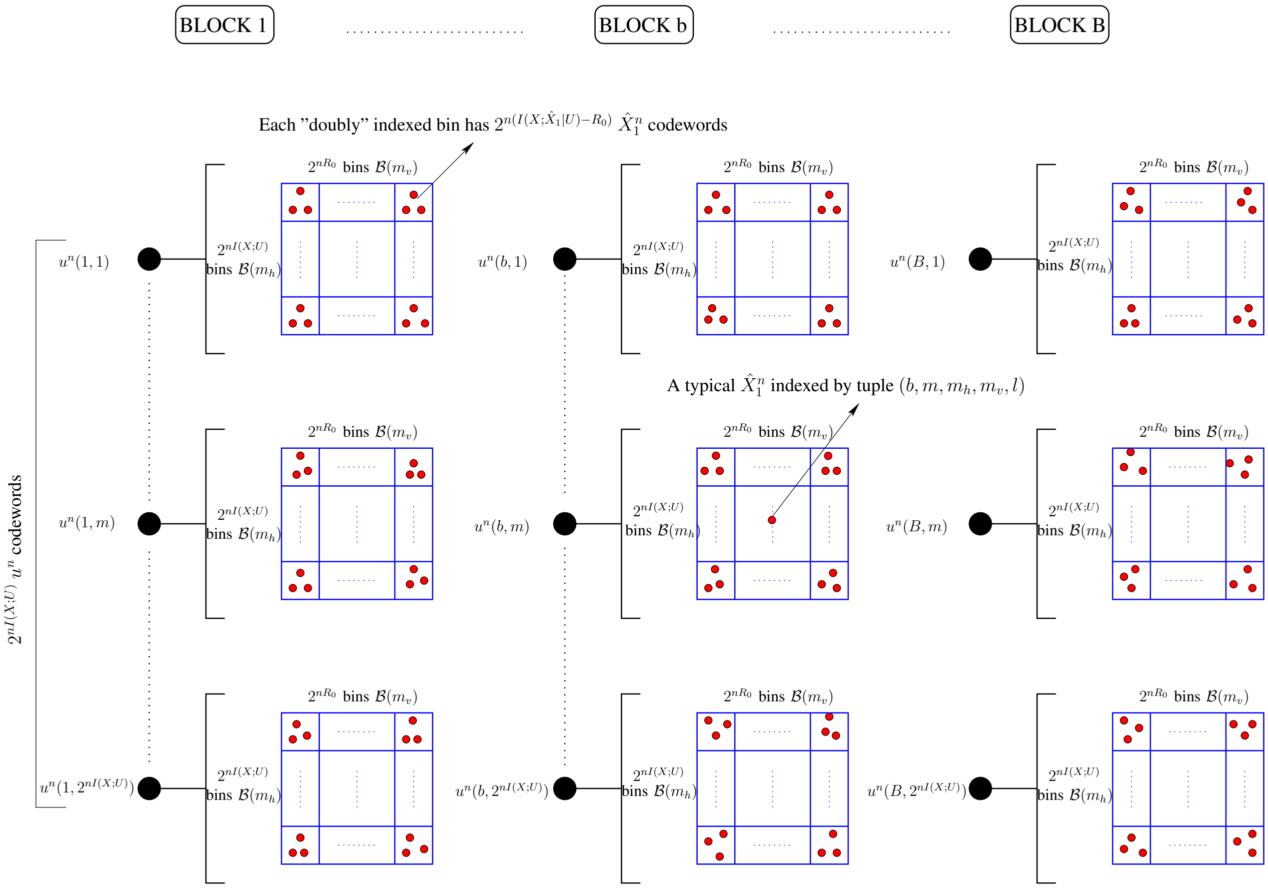

Fig. 10. "Forward Encoding" and "Block Markov Decoding" - achievability scheme for the strictly-causal perfect cribbing.

$\hat{X}_1^n \in \mathcal{B}(m_{v,b})$, the index tuple $(m_{h,b}, l_b)$ is described as R_1 and $m_{v,b}$ is described as R_0.

c) In the last block $b = B$, the encoder knows m_B from encoding procedure in the previous block. Fix $m_{B+1} = 1$. Encoder identifies $U^n(B, m_B)$ from the codebook $\mathcal{C}_U(B)$, looks in the corresponding row or horizontal bin $m_{h,B} = m_{B+1}$ and finds the index tuple $(m_{v,B}, l_B)$ such that $(\hat{X}_1^n(B, m_B, m_{h,B}, m_{v,B}, l_B), X^n(B), U^n(B, m_B)) \in \mathcal{T}_\epsilon^n$. As found $\hat{X}_1^n \in \mathcal{B}(m_{v,B})$, the index tuple $(m_{h,B}, l_B)$ is described as R_1 and $m_{v,B}$ is described as R_0.

Hence the encoding has a "Forward Encoding" interpretation, as we encoded the source sequence of the future block as the row or horizontal bin index of the "doubly-binned" codebook in the present

block. As at each block b, R_1 encodes for $(m_{h,b}, l_b)$, thus

$$I(X;U) + I(X;\hat{X}_1|U) - R_0 \leq R_1. \tag{65}$$

3) *Decoding :* Decoding for both the decoders is as follows :

Decoder 1

a) For the first block $b = 1$, Decoder 1 knows $m_1 = 1$, and since it knows the index $(m_{h,1}, m_{v,1}, l_1)$ it identifies $\hat{X}_1^n(1) = \hat{X}_1^n(1, m_1, m_{h,1}, m_{v,1}, l_1)$ as its source estimate for the first block.
b) For the block b $(\in [2 : B])$, Decoder 1 knows m_b from the index sent by the encoder in the $(b-1)$ block (as $m_{h,b-1} = m_b$) and since it knows the index $(m_{h,b}, m_{v,b}, l_b)$ for the current block, it identifies $\hat{X}_1^n(b) = \hat{X}_1^n(b, m_b, m_{h,b}, m_{v,b}, l_b)$, as its source estimate.

Decoder 2

a) For the first block $b = 1$, Decoder 2 assumes $\hat{m}_1 = 1$ and generates its estimate $\hat{X}_2^n(1) = f(U^n(1, \hat{m}_1))$.
b) For the block b $(\in [2 : B])$, Decoder 2 has already estimated $\hat{m}_{b-1}$ in $b-1$ block. It also knows $\hat{X}_1^n(b-1)$(because of strictly causal cribbing) and $m_{v,b-1}$ through R_0. It then looks into the vertical bin with index $m_{v,b-1}$ in the codebook corresponding to the codeword $U^n(b-1, \hat{m}_{b-1})$, and finds a unique row or horizontal bin index $\hat{m}_{h,b-1}$ such that $\hat{X}_1^n(b-1) = \hat{X}_1^n(b-1, \hat{m}_{b-1}, \hat{m}_{h,b-1}, m_{v,b-1}, \tilde{l}_{b-1})$ for some $\tilde{l}_{b-1} \in [1 : 2^{n(I(X;\hat{X}_1|U)-R_0)}]$. But note that estimating $\hat{m}_{h,b-1}$ is equivalent to estimating $\hat{m}_b$, because of our forward encoding procedure, thus Decoder 2 constructs its source estimate for the block b as $\hat{X}_2^n(b) = f(U^n(b, \hat{m}_b))$.

Decoding has a "Block Markov Decoding" interpretation as we see that the decoding for both decoders relies on what was successfully decoded in the previous block.

4) *Rate Region and Bounding Distortion :* We assume without loss of generality, $d_i(\cdot,\cdot) \leq D_{max} < \infty$, $i = 1, 2$. In the encoding and decoding scheme, m_1 was chosen to be a fixed value, deterministically chosen prior to the compression, agreed upon by both encoders and decoders. Hence, for both the decoders distortion in general will not be met for the first block, however we are generous enough to allow for maximum distortion for the first block, which will eventually have insignificant impact on total distortion as the number of blocks becomes large. Consider the following encoding and

decoding events which will help to bound the distortion at Decoder 1 and Decoder 2. Suppose in block $b-1$ and b, index tuples $(m_{h,b-1}, l_{b-1})$ and $(m_{h,b}, l_b)$ are described by the encoder to the Decoder 1, and that $m_{h,b-1} = m_b$ and $m_{h,b} = m_{b+1}$, $\forall\ b = [2:B]$.

a) *Encoding Events* :

- $$\begin{aligned} \mathcal{E}_{e,1}(b) &= \text{No } U^n \text{ sequence is jointly typical with source in block } b & (66)\\ &= \Big\{(X^n(b), U^n(b,\tilde{m}_b)) \notin \mathcal{T}_\epsilon^n\ \forall\ \tilde{m}_b \in [1:2^{nI(X;U)}]\Big\}, & (67)\end{aligned}$$

  for $b = [2:B]$. By Covering Lemma 3, the probability of this event goes to zero as there are $2^{nI(X;U)}$ U^n codewords Similarly, $P(\mathcal{E}_{e,1}(b+1)) \to 0$. Suppose, $(X^n(b), U^n(b, m_b)) \in \mathcal{T}_\epsilon^n$ and $(X^n(b+1), U^n(b+1, m_{b+1})) \in \mathcal{T}_\epsilon^n$, thus row index in block b is $m_{h,b} = m_{b+1}$.

- $$\begin{aligned} \mathcal{E}_{e,2}(b) &= \text{No } \hat{X}_1^n \text{ sequence is jointly typical with the typical pair } (X,U) \text{ in block } b\\ &= \mathcal{E}^c_{e,1}(b) \cap \mathcal{E}^c_{e,1}(b+1)\\ &\quad \cap \Big\{(\hat{X}_1^n(b, m_b, m_{h,b}, \tilde{m}_{v,b}, \tilde{l}_b), X^n(b), U^n(b, m_b)) \notin \mathcal{T}_\epsilon^n\ \forall \text{ tuples } (\tilde{m}_{v,b}, \tilde{t}_b)\Big\}, \end{aligned} \tag{68}$$

  for $b = [1:B]$, where,

  $$\begin{aligned} \mathcal{E}^c_{e,1}(b) &= \Big\{(X^n(b), U^n(b, m_b)) \in \mathcal{T}_\epsilon^n\Big\} & (69)\\ \mathcal{E}^c_{e,1}(b+1) &= \Big\{(X^n(b+1), U^n(b+1, m_{b+1})) \in \mathcal{T}_\epsilon^n\Big\}. & (70)\end{aligned}$$

  By Covering Lemma 3, this event has vanishing probability as for every row index there are, $2^{nI(X;\hat{X}_1|U)}$ $\hat{X}_1^n$ codewords.

b) *Decoding Events* : Decoder 1 can perfectly construct the $\hat{X}_1^n(b)$ sequences in the block b. Decoder 2 in block b knows $\hat{X}_1^n(b-1)$. For the Decoder 2, for $b \in [2:B]$, assume it has decoded correctly the message, $\hat{m}_{b-1} = m_{b-1}$ in the $b-1$ block and the encoder sends the row index $m_{h,b-1} = m_b$ in the block $b-1$ to Decoder 1. Also Decoder 2 knows $m_{v,b-1}$

through R_0. Decoder 2 needs to find an estimate $\hat{m}_{h,b-1}$, or equivalently an estimate of $\hat{m}_b$ (as $m_{h,b-1} = m_b$). Consider the following events :

- $$\begin{aligned}\mathcal{E}_{d,1} &= \hat{X}_1^n(b-1) \text{ does not lie in row with index, } m_{h,b-1} = m_b \text{ and column index } m_{v,b-1} \\ &= \left\{\hat{X}_1^n(b-1) = \hat{X}_1^n(b-1, m_{b-1}, m_{h,b-1}, m_{v,b-1}, \tilde{l}_{b-1})\right\}, \qquad (71)\end{aligned}$$

  $\forall\ \tilde{l} \in [1 : 2^{n(I(X;\hat{X}_1|\hat{X}_2)-R_0)}]$. But the probability of this event goes to zero, because due to our encoding procedure, $\hat{X}_1^n(b-1) = \hat{X}_1^n(b-1, m_{b-1}, m_{h,b-1}, m_{v,b-1}, l_{b-1})$.

- $$\begin{aligned}\mathcal{E}_{d,2} &= \hat{X}_1^n(b-1) \text{ lies in a row with index, } \hat{m}_{h,b-1} \neq m_b \text{ and column index } m_{v,b-1} \\ &= \left\{\hat{X}_1^n(b-1) = \hat{X}_1^n(b-1, m_{b-1}, \hat{m}_{h,b-1}, m_{v,b-1}, \tilde{l}_{b-1}),\ \hat{m}_{h,b-1} \neq m_b\right\}, \qquad (72)\end{aligned}$$

  for some $\tilde{l} \in [1 : 2^{n(I(X;\hat{X}_1|\hat{X}_2)-R_0)}]$. This event is equivalent to finding $\hat{X}_1^n(b-1)$ corresponding to $U^n(b-1, m_{b-1})$ lying in two different rows or horizontal bins, but with the same column or vertical bin index ($m_{v,b-1}$). The probability of a single $\hat{X}_1^n$ codeword (corresponding to a U^n codeword) occurring repeatedly in two horizontal bins indexed with different row index is $2^{-nH(\hat{X}_1|U)}$, while knowing the column index, total number of $\hat{X}_1^n$ codewords with a particular column index are, $2^{n(I(X;\hat{X}_1,\hat{X}_2)-R_0)}$, so the probability of event $\mathcal{E}_{d,2}$ vanishes so long as,

  $$I(\hat{X}_1; \hat{X}_1, \hat{X}_2) - R_0 < H(\hat{X}_1|U). \qquad (73)$$

Thus consider the event $\mathcal{E}(b) = \mathcal{E}_{e,1}(b) \cup \mathcal{E}_{e,2}(b) \cup \mathcal{E}_{d,1}(b) \cup \mathcal{E}_{d,2}(b)$. We have,

$$P(\mathcal{E}(b)) \leq P(\mathcal{E}_{e,1}(b)) + P(\mathcal{E}_{e,2}(b)) + P(\mathcal{E}_{d,1}(b)) + P(\mathcal{E}_{d,2}(b)), \qquad (74)$$

which vanishes to zero with large n, for each block $b = [2 : B]$, if [from Eq. (65), Eq. (73)], if,

$$\begin{aligned}R_0 + R_1 &\geq I(X; \hat{X}_1, U) \qquad (75)\\ R_0 &\geq \{I(X; \hat{X}_1, U) - H(\hat{X}_1|U)\}^+. \qquad (76)\end{aligned}$$

We will now bound the distortion. The distortion for both the decoders in the first block is bounded above by D_{max}. Consider the block $b = [2:B]$ for Decoder 1,

$$\begin{aligned}
\mathsf{E}\left[d_1(X^n(b), \hat{X}_1^n(b))\right] &= P(\mathcal{E}(b))E\left[d_1(X^n(b), \hat{X}_1^n(b))|\mathcal{E}(b)\right] \\
&\quad + P(\mathcal{E}^c(b))\mathsf{E}\left[d_1(X^n(b), \hat{X}_1^n(b))|\mathcal{E}^c(b)\right] &(77)\\
&\overset{(a)}{\leq} P(\mathcal{E}(b))D_{max} + P(\mathcal{E}^c(b))(1+\epsilon)\mathsf{E}[d_1(X, \hat{X}_1)] &(78)\\
&\leq P(\mathcal{E}(b))D_{max} + P(\mathcal{E}^c(b))D_1, &(79)
\end{aligned}$$

where (a) follows from Typical Average Lemma 2, as given $\mathcal{E}^c(b)$, $(X^n(b), \hat{X}_1^n(b)) \in \mathcal{T}_\epsilon^n$. Thus as $n \to \infty$, $P(\mathcal{E}(b)) \to 0$, hence the distortion is bounded by D_1 in block b. Similarly for Decoder 2,

$$\begin{aligned}
\mathsf{E}\left[d_2(X^n(b), \hat{X}_2^n(b))\right] &= P(\mathcal{E}(b))E\left[d_2(X^n(b), \hat{X}_2^n(b))|\mathcal{E}(b)\right] \\
&\quad + P(\mathcal{E}^c(b))\mathsf{E}\left[d_2(X^n(b), \hat{X}_2^n(b))|\mathcal{E}^c(b)\right] &(80)\\
&\overset{(b)}{\leq} P(\mathcal{E}(b))D_{max} + P(\mathcal{E}^c(b))(1+\epsilon)\mathsf{E}[d_2(X, \hat{X}_2)] &(81)\\
&\leq P(\mathcal{E}(b))D_{max} + P(\mathcal{E}^c(b))D_2, &(82)
\end{aligned}$$

where (b) follows from Typical Average Lemma 2, as given $\mathcal{E}^c(b)$, $(X^n(b), \hat{U}_2^n(b, m_b)) \in \mathcal{T}_\epsilon^n$, and since $\hat{X}_2^n(b) = f(U^n(b, m_b))$, $(X^n(b), \hat{X}_2^n(b)) \in \mathcal{T}_\epsilon^n$. Thus the distortion is bounded by D_2 in block b. The total normalized distortion in B blocks for Decoder 1 and Decoder 2 is bounded above by $\frac{1}{B}D_{max} + \frac{B-1}{B}D_1$ and $\frac{1}{B}D_{max} + \frac{B-1}{B}D_2$ respectively. Proof is completed by letting, $B \to \infty$.

Converse : Converse in this subsection is skipped and follows from the converse of deterministic function cribbing of the next subsection, by the substitution $\hat{Z}_1 = \hat{X}_1$. ■

2) Deterministic Function Cribbing:

Theorem 4. *The rate region $\mathcal{R}(D_1, D_2)$ for the setting in Fig. 11 with deterministic function cribbing (strictly causal) is given as the closure of the set of all the rate tuples (R_0, R_1) such that,*

$$R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2) \tag{83}$$

$$R_0 \geq \{I(X; \hat{Z}_1, \hat{X}_2) - H(\hat{Z}_1|\hat{X}_2)\}^+, \tag{84}$$

for some joint probability distribution $P_X P_{\hat{Z}_1,\hat{X}_2|X} P_{\hat{X}_1|\hat{Z}_1,\hat{X}_2,X}$ *such that* $\mathsf{E}[d_i(X,\hat{X}_i)] \le D_i$, *for* $i = 1,2$.

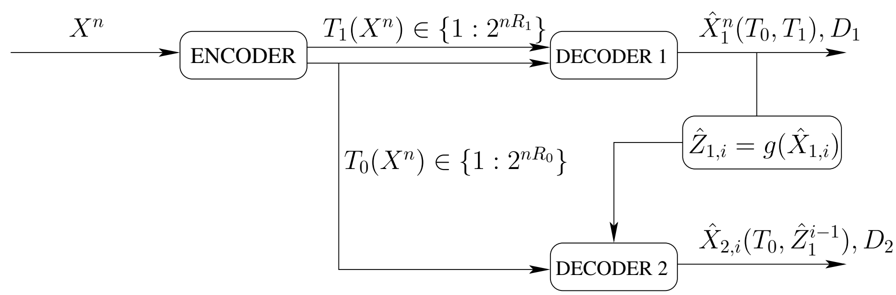

Fig. 11. Successive refinement, with decoders *cooperating* via (deterministic function) *strictly-causal cribbing*.

Proof:

Achievability :

The extension to deterministic function cribbing from perfect cribbing follows similarly to the case of noncausal cribbing in Section III-A2. We omit the details of achievability and describe the key idea. Here also, achievability is first proved with auxiliary random variable U and the following region will be achieved,

$$R_0 + R_1 \ge I(X;\hat{X}_1,U) \tag{85}$$

$$R_0 \ge \{I(X;\hat{Z}_1,U) - H(\hat{Z}_1|U)\}^+, \tag{86}$$

for some joint probability distribution $P_X P_{\hat{Z}_1,U|X} P_{\hat{X}_1|\hat{Z}_1,U,X} \mathbf{1}_{\{\hat{X}_2=f(U)\}}$ such that $\mathsf{E}[d_i(X,\hat{X}_i)] \le D_i$, for $i = 1,2$. The codebook structure remains almost the same, just that instead of (uniformly) binning $\hat{X}_1^n$ into vertical 2^{nR_0} bins, as done in the setting of the previous subsection with perfect cribbing, we bin $\hat{Z}_1^n$ codewords and $\hat{X}_1^n$ codewords are then generated on the top of each $\hat{Z}_1^n$ codewords. Encoding changes accordingly and Decoder 2 tries to infer the row index from the deterministic crib which it obtains from Decoder 1.

Converse : Assume we have a $(2^{nR_0}, 2^{nR_1}, n)$ distortion code (as per Definition 4) such that (R_0,R_1,D_1,D_2) tuple is feasible (as per Definition 2). Denote $T_1 = f_{1,n}(X^n)$ and $T_0 = f_{2,n}(X^n)$.

Identify the auxiliary random variable $U_i = (T_0, \hat{Z}_1^{i-1})$:

$$
\begin{aligned}
H(\hat{Z}_1^n, T_0) &\geq I(X^n; \hat{Z}_1^n, T_0) && (87)\\
&= \sum_{i=1}^{n} I(X_i; \hat{Z}_1^n, T_0 | X^{i-1}) && (88)\\
&\stackrel{(a)}{=} \sum_{i=1}^{n} I(X_i; \hat{Z}_1^n, T_0, X^{i-1}) && (89)\\
&\geq \sum_{i=1}^{n} I(X_i; \hat{Z}_1^i, T_0) && (90)\\
&= \sum_{i=1}^{n} I(X_i; \hat{Z}_{1,i}, U_i) && (91)\\
&\geq nI(X_Q; \hat{Z}_{1,Q}, U_Q) && (92)\\
H(\hat{Z}_1^n, T_0) &= \sum_{i=1}^{n} H(\hat{Z}_{i,1} | T_0, \hat{Z}_1^{i-1}) + H(T_0) && (93)\\
&\leq \sum_{i=1}^{n} H(\hat{Z}_{i,1} | U_i) + nR_0 && (94)\\
&\leq nH(\hat{Z}_{1,Q} | U_Q) + nR_0, && (95)
\end{aligned}
$$

where (a) follows from the independence of X_i with X^{i-1} and $Q \in [1 : n]$ is similarly defined an independent (of source) uniformly distributed time sharing random variable. As argued in previous subsection of perfect cribbing, we lower bound $n(R_0 + R_1)$ with $nI(X; \hat{X}_{1,Q}, U_Q)$. Note that as $\hat{X}_{2,Q} = f(U_Q)$, for some function f. Lastly we bound the distortion for both decoders as we did in previous section and note that the joint distribution of $(X_Q, \hat{X}_{1,Q}, \hat{X}_{2,Q})$ is the same as $(X, \hat{X}_1, \hat{X}_2)$ to derive the rate region with auxiliary random variable. It is easy to see that in inequalities (91) and (94), we can replace U_i with $\hat{X}_{2,i}$ and this helps to provide converse for the region without auxiliary random variable provided in the theorem. ■

C. Causal Cribbing

1) Perfect Cribbing:

Theorem 5. *The rate region $\mathcal{R}(D_1, D_2)$ for the setting in Fig. 9 with perfect causal cribbing that is $\hat{X}_{2,i}$*

is a function of $(T_0, \hat{X}_1^i)$*, is given as the closure of the set of all the rate tuples* (R_0, R_1) *such that,*

$$R_0 + R_1 \geq I(X; \hat{X}_1, U) \tag{96}$$

$$R_0 \geq \{I(X; \hat{X}_1, U) - H(\hat{X}_1|U)\}^+, \tag{97}$$

for some joint probability distribution $P_{X,\hat{X}_1,U}\mathbf{1}_{\{\hat{X}_2=f(U,\hat{X}_1)\}}$ *such that* $\mathsf{E}[d_i(X, \hat{X}_i)] \leq D_i$*, for* $i = 1, 2$ *and* $|\mathcal{U}| \leq |\mathcal{X}|\,|\mathcal{X}_1| + 4$*.*

Proof: The achievability remains the same as in strictly causal cribbing, in terms of encoding and decoding operations at Decoder 1. For Decoder 2, the only change is in constructing $\hat{X}_2^n(b)$ for each block, which in this case is constructed as, $\hat{X}_{2,i}^n(b) = f(U_i(b, m_b), \hat{X}_{1,i})$. The steps in the converse are exactly the same as in the strictly causal cribbing case, except that this time we identify $\hat{X}_{2,Q} = f(U_Q, \hat{X}_{1,Q})$. The cardinality bounds on $\mathcal{U}$ follow standard arguments as in [21] : $\mathcal{U}$ should have $|\mathcal{X}|\,|\mathcal{X}_1| - 1$ elements to preserve the joint probability distribution $P_{X,\hat{X}_1}$, one element to preserve the markov chain, $(X, \hat{X}_1) - U - \hat{X}_2$, two elements to preserve the mutual information quantities, $I(X; \hat{X}_1, U)$ and $\{I(X; \hat{X}_1, U) - H(\hat{X}_1|U)\}^+$ and finally two more elements to preserve the distortion constraints. ■

2) Deterministic Function Cribbing:

Theorem 6. *The rate region* $\mathcal{R}(D_1, D_2)$ *for the setting in Fig. 11 with deterministic function cribbing but with causal cribbing, that is,* $\hat{X}_{2,i}$ *is a function of* $(T_0, \hat{X}_1^i)$*, is given as the closure of the set of all the rate tuples* (R_0, R_1) *such that,*

$$R_0 + R_1 \geq I(X; \hat{X}_1, U) \tag{98}$$

$$R_0 \geq \{I(X; \hat{Z}_1, U) - H(\hat{Z}_1|U)\}^+, \tag{99}$$

for some joint probability distribution $P_X P_{\hat{Z}_1,U|X} P_{\hat{X}_1|\hat{Z}_1,U,X}\mathbf{1}_{\{\hat{X}_2=f(U,\hat{Z}_1)\}}$ *such that* $\mathsf{E}[d_i(X, \hat{X}_i)] \leq D_i$*, for* $i = 1, 2$ *and* $|\mathcal{U}| \leq |\mathcal{X}|\,|\mathcal{X}_1| + 4$*.*

Proof: The achievability remains the same as in strictly causal deterministic function cribbing, in terms of encoding operation and decoding operation at Decoder 1. For the Decoder 2, only change is in constructing $\hat{X}_2^n(b)$ for each block, it is constructed as, $\hat{X}_{2,i}^n(b) = f(U_i(b, m_b), \hat{Z}_{1,i})$. The steps in converse are exactly the same as in strictly causal cribbing case except that we identify, $\hat{X}_{2,Q} = f(U_Q, \hat{Z}_{1,Q})$. ■

IV. Special Cases

In this section, we study some special cases of our setting and also compute certain numerical examples.

A. The Case $R_0 = 0$

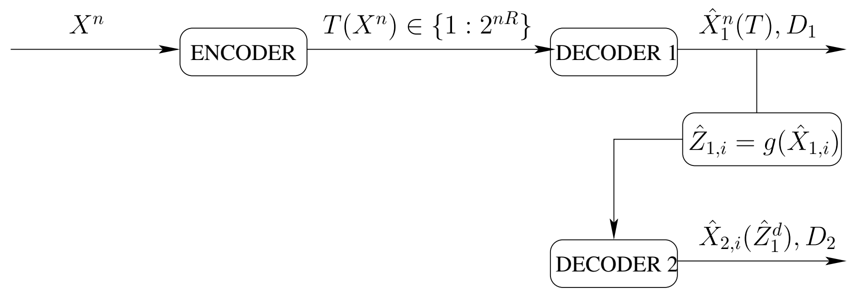

Fig. 12. Special case of successive refinement with cribbing decoders, when the common rate is zero. Here again $d = n$, $d = i-1$ and $d = i$ respectively stand for non-causal, strictly-causal and causal cribbing.

One special yet important case of the setting studied in previous sections, is that when $R_0 = 0$ as shown in Fig. 12. Here the encoder describes the source to only Decoder 1, while Decoder 2 attempts to find the reconstruction of the source within some distortion via *cribbing* reconstruction symbols of the Decoder 1, non-causally, causally or strictly causally. Table II provides the minimum achievable rate ($R = R(D_1, D_2)$) for various cases, derived when $R_0 = 0$, using Theorem 1 through 6. Distortion constraints are omitted for brevity.

B. Null g function

Our expressions reduce to the successive refinement rate region (cf. Equitz and Cover [18]), when g is a trivial function. To see this consider rate region for non-causal cribbing with deterministic cribbing (cf. Theorem 2), as given below,

$$R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2) \quad (100)$$

$$R_0 \geq \{I(X; \hat{Z}_1, \hat{X}_2) - H(\hat{Z}_1)\}^+, \quad (101)$$

$R(D_1, D_2)$	Non-Causal ($d = n$)	Strictly-Causal ($d = i-1$)	Causal ($d = i-1$)
Deterministic Function Cribbing	$\min I(X; \hat{X}_1, \hat{X}_2)$ s.t. $I(X; \hat{Z}_1, \hat{X}_2) \leq H(\hat{Z}_1)$ *(p.m.f.)* : $P(X, \hat{X}_1, \hat{X}_2) \times$ $\mathbf{1}_{\{\hat{Z}_1 = f(\hat{X}_1)\}}$	$\min I(X; \hat{X}_1, \hat{X}_2)$ s.t. $I(X; \hat{Z}_1, \hat{X}_2) \leq H(\hat{Z}_1 \vert \hat{X}_2)$ *(p.m.f.)* : $P(X, \hat{X}_1, \hat{X}_2)$ $\mathbf{1}_{\{\hat{Z}_1 = f(\hat{X}_1)\}}$	$\min I(X; \hat{X}_1, U)$ s.t. $I(X; \hat{Z}_1, U) \leq H(\hat{Z}_1 \vert U)$ *(p.m.f.)* : $P(X, \hat{X}_1, U) \times$ $\mathbf{1}_{\{\hat{Z}_1 = f(\hat{X}_1), \hat{X}_2 = f(\hat{Z}_1, U)\}}$

TABLE II
RESULTS FOR THE SUCCESSIVE REFINEMENT WITH CRIBBING DECODERS, WHEN COMMON RATE, $R_0 = 0$.

If g is null, $\hat{Z}_1$ is constant and hence the region reduces to,

$$R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2) \tag{102}$$

$$R_0 \geq I(X; \hat{X}_2), \tag{103}$$

for some joint probability distribution $P_{X,\hat{X}_1,\hat{X}_2}$ such that $\mathsf{E}[d_i(X, \hat{X}_i)] \leq D_i$, for $i = 1, 2$, which is also derived in Equitz and Cover [18].

C. Numerical Examples

We provide an example illustrating the rate regions of non-causal and strictly causal cribbing. Along with them, the region without cribbing is also compared. The rate regions for these three cases from the theorems in the paper are shown in the Table III. Distortion constraints are omitted for brevity.

We plot for a specific example (cf. setting in Fig. 4 with perfect cribbing) with a bernoulli source $X \sim \text{Bern}(0.5)$, binary reconstruction alphabets and hamming distortion. We consider a particular distortion tuple (D_1, D_2). Due to symmetry of the source, for the optimal distribution, it is easy to argue that, $P_{\hat{X}_1,\hat{X}_2|X}(\hat{x}_1, \hat{x}_2|x) = P_{\hat{X}_1,\hat{X}_2|X}(\overline{\hat{x}_1}, \overline{\hat{x}_2}|\overline{x})$, where $\overline{x}$ stands for complement of x. Thus all the expressions can be written in terms of variables $p_1 = P_{\hat{X}_1,\hat{X}_2|X}(0,0|0)$, $p_2 = P_{\hat{X}_1,\hat{X}_2|X}(0,1|0)$, $p_3 = P_{\hat{X}_1,\hat{X}_2|X}(1,0|0)$ and $p_4 = P_{\hat{X}_1,\hat{X}_2|X}(1,1|0)$, $p_4 = 1 - p_1 - p_2 - p_3$. However it is also easy to see that the distortion constraints are satisfied with equality, otherwise one can reduce the rate region slightly and still be under

Non-Causal Cribbing	Strictly-Causal Cribbing	No Cribbing
$R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2)$ $R_0 \geq \{I(X; \hat{X}_1, \hat{X}_2) - H(\hat{X}_1)\}^+$ *(p.m.f.)* $: P(X, \hat{X}_1, \hat{X}_2)$	$R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2)$ $R_0 \geq \{I(X; \hat{X}_1, \hat{X}_2) - H(\hat{X}_1\|\hat{X}_2)\}^+$ *(p.m.f.)* $: P(X, \hat{X}_1, \hat{X}_2)$	$R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2)$ $R_0 \geq I(X; \hat{X}_2)$ *(p.m.f.)* $: P(X, \hat{X}_1, \hat{X}_2)$

TABLE III
COMPARING RATE REGIONS FOR THE EXAMPLE CONSIDERED, FOR NON-CAUSAL CRIBBING, STRICTLY CAUSAL CRIBBING AND NO CRIBBING.

distortion constraint. The distortion constraints thus yield,

$$\mathsf{E}[d(X, \hat{X}_1)] = p_4 + p_3 = D_1 \quad (104)$$

$$\mathsf{E}[d(X, \hat{X}_2)] = p_2 + p_4 = D_2, \quad (105)$$

which implies, $p_2 = 1 - D_1 - p_1$, $p_3 = 1 - D_2 - p_1$, $p_4 = p_1 + D_1 + D_2 - 1$. Thus the equivalent probability distribution space over which the closure of rate regions is evaluated (such that distortion is satisfied) is equivalent to, $\mathcal{P} = \{p_1 \in [1 - D_1 - D_2, \min\{1 - D_1, 1 - D_2, 2 - D_1 - D_2\}], p_2 = 1 - D_1 - p_1, p_3 = 1 - D_2 - p_1, p_4 = p_1 + D_1 + D_2 - 1\}$. The various entropy and mutual information expressions appearing in the rate regions of non-causal, strictly causal and no cribbing (cf. Table III) can then be expressed as,

$$I(X; \hat{X}_1, \hat{X}_2) = H_2\Big(\Big[\frac{p_1 + p_4}{2}\ \frac{p_2 + p_3}{2}\ \frac{p_2 + p_3}{2}\ \frac{p_1 + p_4}{2}\Big]\Big) - H_2([p_1\ p_2\ p_3\ p_4]) \quad (106)$$

$$H(\hat{X}_1) = 1 \quad (107)$$

$$H(\hat{X}_1|\hat{X}_2) = H_2\Big(\Big[p_1 + p_4\ p_2 + p_3\Big]\Big) \quad (108)$$

$$I(X; \hat{X}_2) = 1 - H_2\Big(\Big[p_1 + p_3\ p_2 + p_4\Big]\Big), \quad (109)$$

where $H_2(\cdot)$ stands for the binary entropy of the probability vector. Note the only variable of optimization is effectively p_1. Fig. 13 shows the rate regions for $(D_1, D_2) = (0.05, 0.1)$. Note that the region for no cribbing is smaller than that of strictly causal cribbing which is smaller than that of non-causal cribbing,

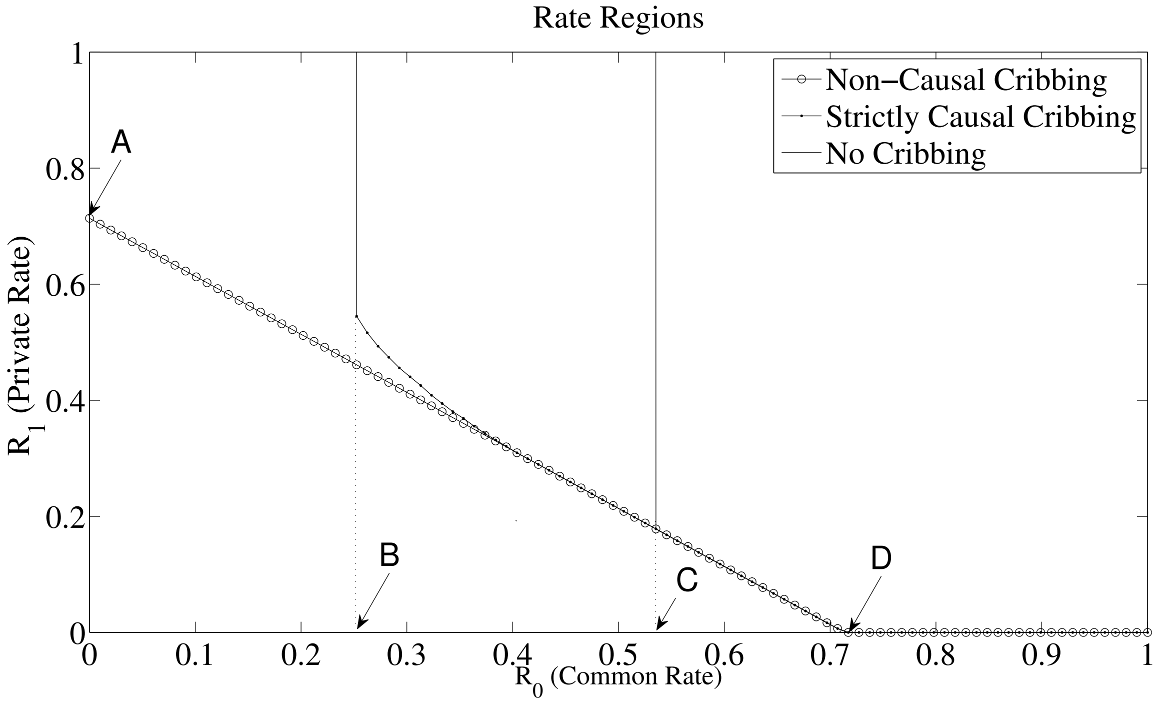

Fig. 13. Rate regions for non-causal, strictly causal and no cribbing in successive refinement setting of Fig. 4. Source is Bern(0.5) and $(D_1, D_2) = (0.05, 0.1)$. The curve is tradeoff curve between R_1 and R_0 and the rate regions lie to the right of the respective tradeoff curves.

as expected. We can also analytically compute the expression of corner points A,B,C,D in Fig. 13. Let $h_2(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)\ \forall\ \alpha \in [0,1]$. Consider independent bernoulli random variables $Z_{D_1} \sim$ Bern(D_1) and $Z_{D_2} \sim$ Bern(D_2). R_0 for point D is evaluated by putting $R_1 = 0$ in rate region for non-causal cribbing and this equals $\min_{\mathcal{P}} I(X; \hat{X}_1, \hat{X}_2)$. We will now show that $\min_{\mathcal{P}} I(X; \hat{X}_1, \hat{X}_2) = 1 - h_2(D_1)$. Consider, $\min_{\mathcal{P}} I(X; \hat{X}_1, \hat{X}_2) \geq \min_{\mathcal{P}} I(X; \hat{X}_1) \geq \min_{\mathcal{P}}(1 - H(\hat{X}_1|X)) \geq 1 - h_2(D_1)$, where the last two inequalities follow respectively as $\hat{X}_1$ is Bern(0.5) and that D_1 is the hamming distortion between $\hat{X}_1$ and X. As $D_2 > D_1$, this lower bound is indeed achieved if $\hat{X}_2 = \hat{X}_1 = X \oplus Z_{D_1}$. Similarly for point A, R_1 is obtained by substituting $R_0 = 0$ in the expression of rate region for non-causal cribbing and this again equals $1 - h_2(D_1)$. R_0 corresponding to points B and C is obtained by putting $R_1 = \infty$ in the expressions of rate regions of strictly-causal and no cribbing. Let us first consider point B and observe that R_0 equals $\min_{\mathcal{P}}\{I(X; \hat{X}_1, \hat{X}_2) - H(\hat{X}_1|\hat{X}_2)\}^+$. We show that this equals $1 - h_2(D_1) - h_2(D_2)$. To see this, consider, $\min_{\mathcal{P}}\{I(X; \hat{X}_1, \hat{X}_2) - H(\hat{X}_1|\hat{X}_2)\}^+ = \min_{\mathcal{P}}\{H(\hat{X}_2) - H(\hat{X}_1, \hat{X}_2|X)\}^+ \geq \min_{\mathcal{P}}\{1 -$

$H(\hat{X}_1|X) - H(\hat{X}_1|X)\}^+$, where the last inequality follows as $\hat{X}_2$ is Bern(0.5). Since $\hat{X}_1$ and $\hat{X}_2$ are within hamming distortion D_1 and D_2 to X respectively, we have $\min_{\mathcal{P}}\{I(X;\hat{X}_1,\hat{X}_2) - H(\hat{X}_1|\hat{X}_2)\}^+ \geq 1 - h_2(D_1) - h_2(D_2)$, where the equality holds for $\hat{X}_1 = X \oplus Z_{D_1}$ and $\hat{X}_2 = X + Z_{D_2}$. Similarly for point C, it can be shown R_0 equals $\min_{\mathcal{P}} I(X;\hat{X}_2) = 1 - h_2(D_2)$.

V. Dual Channel Coding Setting

In this section we establish duality between cribbing decoders in the successive refinement problem and cribbing encoders in the MAC problem with a common message. The duality between rate-distortion and channel capacity was first mentioned by Shannon, [22] and was further developed for the case of side information by Pradhan et. al., [23] and by Chiang and Cover, [24]. Additional duality has been shown by Yu, [25] for a class of broadcast channels and multiterminal source coding problems, and by Shirazi et. al., [26] for the case of increased partial side information. The duality between source and channel coding with action dependent side information was shown in Kittichokechai et al. in [27]. Recently, Gupta and Verdú, [28] have shown operational duality between the codes of source coding and of channel coding with side information.

To make the notion of duality clearer and sharper, we consider coordination problems in source coding [29] and for channel coding we consider a new kind of problems which we refer to as channel coding with restricted code distribution. In the (weak) coordination problem [29] the goal is to generate a joint typical distribution of the sources and the reconstruction (or actions) rather than a distortion constraint between the source and its reconstruction. Similarly, we define a channel coding problem where the code is restricted to a specific type. The achievability proofs for coordination and channel capacity with restricted code distribution are the same as that of rate-distortion and channel capacity, respectively, since the codes in all achievability proofs are generated randomly with specific distribution. The converse is also similar except in the last step where we need to justify the constraint of having a code with a specific type. For this purpose we invoke [29, Property 2] that is stated as follows :

Lemma 5 (Equivalence of type and time-mixed variables [29])**.** *For a collection of random sequences X^n, Y^n, and Z^n, the expected joint type $\mathbf{E}P_{X^n,Y^n,Z^n}$ is equal to the joint distribution of the time-mixed variables (X_Q, Y_Q, Z_Q), where Q is a r.v. uniformly distributed over the integers $\{1, 2, 3, ..., n\}$ and independent of (X^n, Y^n, Z^n).*

The duality principle between source coding and channel coding with cribbing appears later in Table IV. According to those principles, the standard successive refinement source coding problem which was introduced in [18] is dual to the MAC with one common message and one private message [30]. Furthermore, the successive refinement source coding with cribbing decoders is dual to the MAC with one common message and one private message and cribbing encoders. To show the duality, let us investigate the capacity of the MAC with common message and cribbing encoders and compare it to the achievable region of the successive refinement problem with cribbing.

A. MAC with cribbing encoders and a common message

We consider here the problem of MAC with partial cribbing encoders where there is one private message $m_1 \in \{1, 2, ..., 2^{nR_1}\}$ known to Encoder 1 and one common message $m_0 \in \{1, 2, ..., 2^{nR_0}\}$ known to both encoders that needs to be sent to the decoder, as shown in Fig. 14 . We assume that Encoder 2 cribs the signal from Encoder 1, namely, Encoder 2 observes a deterministic function of the output of Encoder 1. We consider here three cases, noncausal, strictly-causal and causal cribbing and we show in the next subsection their duality to the successive refinement problem.

Definition 4. *A ($2^{nR_0}, 2^{nR_1}, n, P(x_1, x_2)$) partial cribbing MAC, with one private and one common message and a code restricted to a distribution $P(x_1, x_2)$, has,*

1) *Encoder 1,* $g_1 : \{1, ..., 2^{nR_0}\} \times \{1, ..., 2^{nR_1}\} \to \mathcal{X}_1^n$.
2) *Encoder 2,* $\forall\ i = 1, ..., n$. *(depending on d in Fig. 4, the decoder mapping changes as below),*

$$g_{2,i}^{nc} \;:\; \{1, ..., 2^{nR_0}\} \times \mathcal{Z}_1^n \to \mathcal{X}_2 \quad \textit{non-causal cribbing, } d = n \tag{110}$$

$$g_{2,i}^{sc} \;:\; \{1, ..., 2^{nR_0}\} \times \mathcal{Z}_1^{i-1} \to \mathcal{X}_2 \quad \textit{strictly-causal cribbing, } d = i-1, \tag{111}$$

$$g_{2,i}^{c} \;:\; \{1, ..., 2^{nR_0}\} \times \mathcal{Z}_1^{i} \to \mathcal{X}_2 \quad \textit{causal cribbing, } d = i. \tag{112}$$

3) *Decoder,* $f : \mathcal{Y}^n \to \{1, ..., 2^{nR_0}\} \times \{1, ..., 2^{nR_1}\}$.

An error occurs if the one of the messages was incorrectly decoded or if the joint type of the output and input to the channel deviates from the required one. Hence, the probability of error is defined for any

integer n and $\delta > 0$ such as

$$Pe^{(n),\delta} = \Pr\left\{(\hat{M}_0(Y^n), \hat{M}_1(Y^n)) \neq (M_0, M_1) \text{ AND } \|P_{X_1^n, X_2^n, Y^n}(x, y, z) - P(x_1, x_2)P(y|x_1, x_2)\|_{TV} \geq \delta\right\}, \tag{113}$$

where $P_{X_1^n, X_2^n, Y^n}(x, y, z)$ is the joint type of the input and output of the channel and $\|\cdot\|_{TV}$ is the total variation between two probability mass functions, i.e., half the L_1 distance between them, given by

$$\|p(x,y,z) - q(x,y,z)\|_{TV} \triangleq \frac{1}{2}\sum_{x,y,z} |p(x,y,z) - q(x,y,z)|.$$

A pair rate (R_0, R_1) is achievable if for any $\delta > 0$ there exists a sequence of codes such that $Pe^{(n),\delta} \to 0$ as $n \to \infty$. The capacity region is defined in the standard way for MAC as in [31, Chapter 15.3], as the union of all achievable rate pairs. Let us define three regions $\mathcal{R}^{nc}, \mathcal{R}^{sc}$ and $\mathcal{R}^c$, which correspond to noncausal, strictly-causal, and causal cases.

$$\mathcal{R}^{nc}(P) \triangleq \begin{cases} R_1 \leq I(Y; X_1|X_2, Z_1) + H(Z_1) \\ R_0 + R_1 \leq I(Y; X_1, X_2), \end{cases} \tag{114}$$

$$\mathcal{R}^{sc}(P) \triangleq \begin{cases} R_1 \leq I(Y; X_1|X_2, Z_1) + H(Z_1|X_2) \\ R_0 + R_1 \leq I(Y; X_1, X_2). \end{cases} \tag{115}$$

$$\mathcal{R}^{c}(P) \triangleq \bigcup_{P(u|x_1)\mathbf{1}_{x_2=f(u,z_1)}} \begin{cases} R_1 \leq I(Y; X_1|U, Z_1) + H(Z_1|U) \\ R_0 + R_1 \leq I(Y; X_1, U), \end{cases} \tag{116}$$

where the union is over joint distributions that preserve the constraint $P(x_1, x_2)$. Since $x_2 = f(u, z_1)$, note that $I(Y; X_1, U) = I(Y; X_1, X_2)$. The next theorem states that the regions defined above, $\mathcal{R}^{nc}(P), \mathcal{R}^{sc}(P)$ and $\mathcal{R}^c(P)$ are the respective capacity regions.

Theorem 7 (MAC with common message and cribbing encoders)**.** *The capacity regions of MAC with common message, restricted code distribution $P(x_1, x_2)$ and non-causal, strictly-causal and causal cribbing that is depicted in Fig. 14 are $\mathcal{R}^{nc}(P), \mathcal{R}^{sc}(P)$ and $\mathcal{R}^c(P)$, respectively.*

The achievability and the converse proof of the theorem is presented in the Appendix. In the coding scheme of the achievability proof, we use block Markov coding, backward decoding and rate splitting

similar to the techniques used in Willems and Van der Muelen [1] and Permuter and Asnani [2]. The converse uses the standard Fano's inequalities and the identification of an auxiliary random variable.

B. Duality results between successive refinement and MAC with a common message

SOURCE CODING	CHANNEL CODING
Source encoder	Channel decoder
Encoder input X_i	Decoder input Y_i
Encoder output $M \in \{1, 2, .., 2^{nR}\}$	Decoder output $M \in \{1, 2, .., 2^{nR}\}$
Encoder function $f : \mathcal{X}^n \mapsto \{1, 2, ..., 2^{nR}\}$	Decoder function $f : \mathcal{X}^n \mapsto \{1, 2, .., 2^{nR}\}$
Source decoder input $M \in \{1, 2, .., 2^{nR}\}$	Channel encoder input $M \in \{1, 2, .., 2^{nR}\}$
Decoder output $\hat{X}^n$	Encoder output X^n
Cribbing decoders $\hat{Z}_i(\hat{X}_i)$	Cribbing encoders $Z_i(X_i)$
Noncausal cribbing decoder $f_i : \{1, 2, ..., 2^{nR}\} \times \hat{Z}^n \mapsto \hat{X}_i$	Noncausal cribbing encoder $f_i : \{1, 2, ..., 2^{nR}\} \times Z^n \mapsto X_i$
Strictly-causal cribbing decoder $f_i : \{1, 2, ..., 2^{nR}\} \times \hat{Z}^{i-1} \mapsto \hat{X}_i$	Strictly-causal cribbing encoder $f_i : \{1, 2, ..., 2^{nR}\} \times Z^{i-1} \mapsto X_i$
Causal cribbing decoder $f_i : \{1, 2, ..., 2^{nR}\} \times \hat{Z}^i \mapsto \hat{X}_i$	Causal cribbing encoder $f_i : \{1, 2, ..., 2^{nR}\} \times Z^i \mapsto X_i$
Auxiliary r.v. U	Auxiliary r.v. U
Constraint $P(x, \hat{x}_1, \hat{x}_2)$, $P(x)$ is fixed	Constraint $P(y, x_1, x_2)$, $P(y\|x_1, x_2)$ is fixed
Joint distribution $P(x, \hat{x}_1, \hat{x}_2, u)$	Joint distribution $P(y, x_1, x_2, u)$

TABLE IV
PRINCIPLES OF DUALITY BETWEEN SOURCE CODING AND CHANNEL CODING

Now that we have the capacity regions of the MAC with common message and of successive refinement we explore the duality of the regions. From a first glance at the regions of MAC with common message and of successive refinement, their duality may go unnoticed. However, the corner points of the regions are dual according to the principles presented in Table IV and as seen in Fig. 14.

Tables VII-VI presents the corner points of the capacity region of the MAC with partial cribbing and common message and compare them to the corner points of the successive refinement (SR) rate region

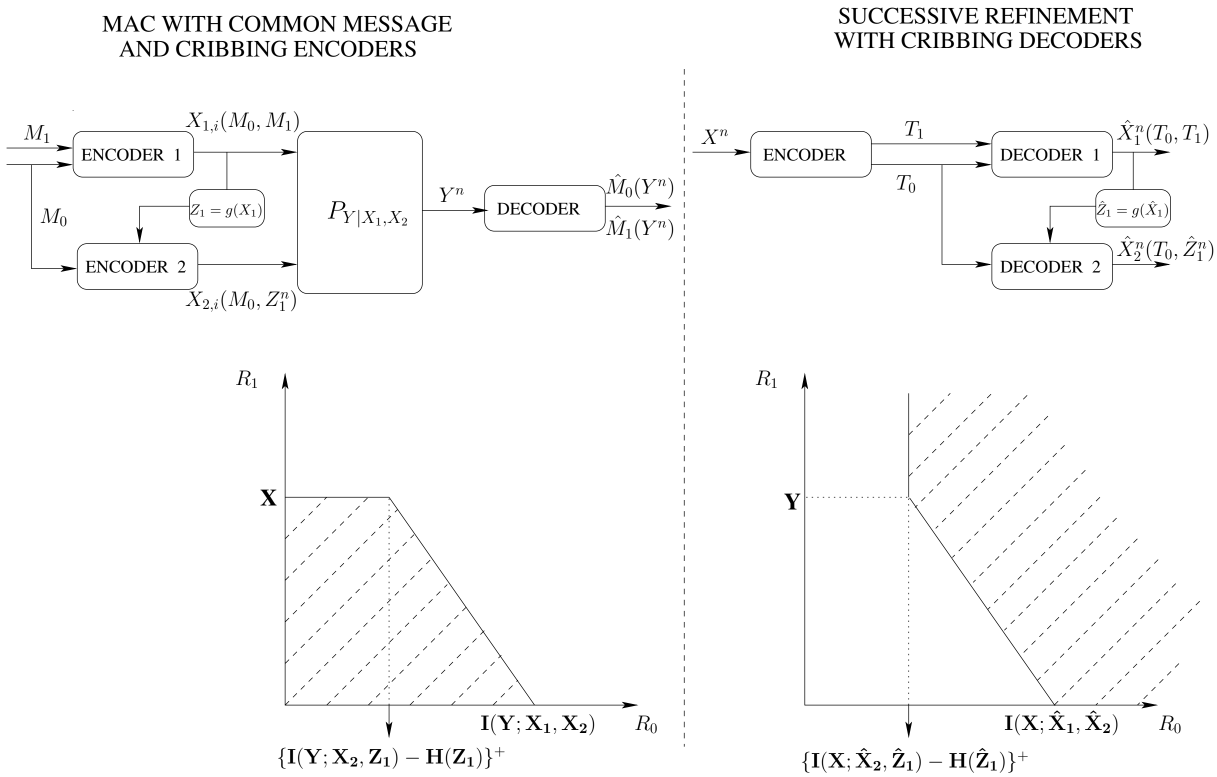

Fig. 14. Duality between the cribbing decoders in successive refinement problem and the cribbing encoders in the MAC problem with a common message, non-causal case. Table IV represents how the expression of rate and capacity regions of the two problems are related. In the figure, for a fixed joint probability distribution, we plot the rate and capacity regions, and we observe that the corner points are dual to each other. Point $\mathbf{Y}$ corresponds to $(R_0, R_1) = (0, I(X;\hat{X}_1,\hat{X}_2) - \{I(X;\hat{X}_2,Z_1) - H(Z_1)\}^+)$ and Point $\mathbf{X}$ corresponds to $(R_0, R_1) = (0, I(Y;X_1,X_2) - \{I(Y;X_2,Z_1) - H(Z_1)\}^+)$.

	Corner points (R_0, R_1) of the noncausal $(d = n)$
MAC Eq. (114)	$(I(Y;X_1,X_2), 0)$ $(\{I(Y;X_2,Z_1) - H(Z_1)\}^+, I(Y;X_1,X_2) - \{I(Y;X_2,Z_1) - H(Z_1)\}^+)$
SR Theorem 2	$(I(X;\hat{X}_1,\hat{X}_2), 0)$ $(\{I(X;\hat{X}_2,Z_1) - H(Z_1)\}^+, I(X;\hat{X}_1,\hat{X}_2) - \{I(X;\hat{X}_2,Z_1) - H(Z_1)\}^+)$

TABLE V
THE CORNER POINTS OF THE NONCAUSAL CASE.

with partial cribbing encoders. Note that applying the dual rules $X_1 \leftrightarrow \hat{X}_1$, $X_2 \leftrightarrow \hat{X}_2$, $Y \leftrightarrow X$, and $\geq\leftrightarrow\leq$, we obtain duality between the corner points of the capacity region of MAC with common message and the rate region of the successive refinement setting.

	Corner points (R, R_1) of the strictly causal case $(d = i - 1)$
MAC Eq. (115)	$(I(Y; X_1, X_2), 0)$, $(\{I(Y; X_2, Z_1) - H(Z_1\|X_2)\}^+, I(Y; X_1, U) - \{I(Y; X_2, Z_1) - H(Z_1\|X_2)\}^+)$
SR Theorem 4	$(I(X; \hat{X}_1, \hat{X}_2, 0)$ $(\{I(X; \hat{X}_2, Z_1) - H(Z_1\|\hat{X}_2)\}^+, I(X; \hat{X}_1, \hat{X}_2) - \{I(X; \hat{X}_2, Z_1) - H(Z_1\|\hat{X}_2)\}^+)$

TABLE VI
THE CORNER PONTS OF THE STRICTLY CAUSAL CASE.

	Corner points (R, R_1) of the causal case $(d = i)$
MAC Eq. (116)	$(I(Y; X_1, U), 0)$, where $X_2 = f(U, Z_1)$ $(\{I(Y; U, Z_1) - H(Z_1\|U)\}^+, I(Y; X_1, U) - \{I(Y; U, Z_1) - H(Z_1\|U)\}^+)$
SR Theorem 6	$(I(X; \hat{X}_1, U, 0)$ where $\hat{X}_2 = f(U, Z_1)$ $(\{I(X; U, Z_1) - H(Z_1\|U)\}^+, I(X; \hat{X}_1, U) - \{I(X; U, Z_1) - H(Z_1\|U)\}^+)$

TABLE VII
THE CORNER POINTS OF THE CAUSAL CASE.

C. Duality between MAC with conferencing encoders and successive refinement with conferencing decoders

In the previous subsection we saw that there is a duality between the problem of MAC with one common message and one private message with cribbing encoders to successive refinement with cribbing decoders. Now we show that the duality also exists if the cooperation between the encoders/decoders is through a limited rate (conferencing) link as shown in Fig. 15.

Theorem 8. *The capacity region of MAC with one common message at rate R_0 known to both encoders, one private message at rate R_1 known to Encoder 1, and a limited rate link from Encoder 1 to Encoder 2 at rate R_{12} with a restricted code distribution $P(x_1, x_2)$ is*

$$\begin{aligned} R_0 + R_1 &\leq I(X_1, X_2; Y) \\ R_1 &\leq I(X_1; Y|X_2) + R_{12}. \end{aligned} \tag{117}$$

This theorem can be proved using the result of conferencing MAC [32] where $C_{21} = \infty$, and choosing $U = X_2$. It is also possible to prove the theorem directly. The achievability part of Theorem 8 follows

easily if $R_1 \leq R_{12}$, then the conferencing link can be used to convey message M_1, thus both the encoders have a common knowledge of both the messages, so $R_0 + R_1 \leq I(X_1, X_2; Y)$ is achievable. If rate $R_{12} \leq R_1$, then the conferencing can be used to increase the common message rate to $R_0 + R_{12}$ and decrease the private message rate to $R_1 - R_{12}$. The converse can be proved using the fact that $nR_1 = H(M_1) = H(M_1|M_0) \leq H(M_{12}|M_0) + H(M_1|M_0, M_{12})$, and then bounding $H(M_{12}|M_0) \leq nR_{12}$ using the fact that the cardinality of M_{12} is $2^{nR_{12}}$ and bounding $H(M_1|M_0, M_{12}) \leq nI(X_{1Q}; Y_Q|X_{2Q})$ using Fano's inequality and the fact that the channel is memoryless.

Finally, one can note a duality between the MAC with one common message and one private message and conferencing encoders to the successive refinement with conferencing decoders. In particular, Table VIII presents the corner points of the achievability regions of the two problems from which the duality rules $X_1 \leftrightarrow \hat{X}_1$, $X_2 \leftrightarrow \hat{X}_2$, $Y \leftrightarrow X$, emerge.

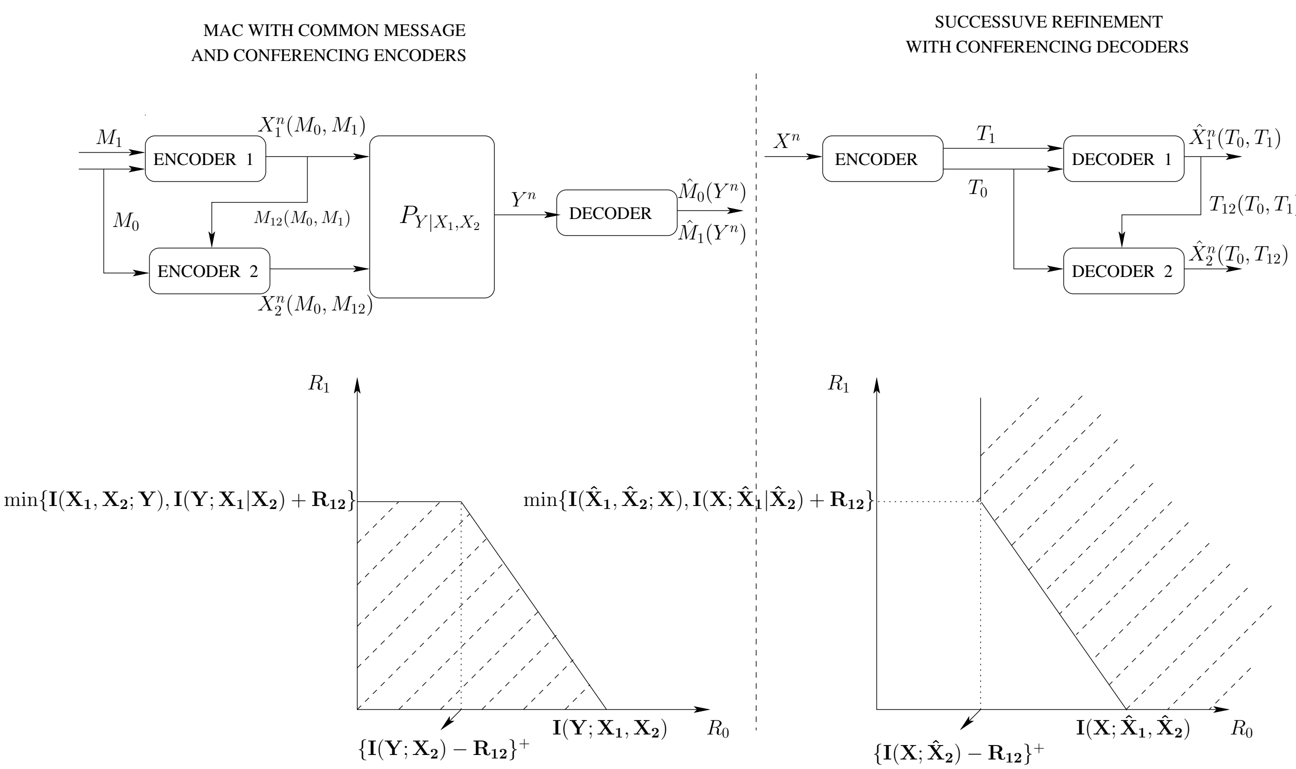

Fig. 15. Duality between the conferencing decoders in successive refinement problem and the conferencing encoders in the MAC problem with a common message, non-causal case. In the figure, for a fixed joint probability distribution, we plot the rate and capacity regions, and we observe that the corner points are dual to each other.

	Corner points (R_0, R_1) of the conferencing case
MAC Theorem 8	$(I(Y;X_1,X_2),0)$ $(\min(I(Y;X_1,X_2),I(Y;X_1\|X_2)+R_{12}),\{I(Y;X_2)-R_{12}\}^+)$
SR Eq. (1)-(2)	$(I(X;\hat{X}_1,\hat{X}_1),0)$ $(\min(I(X;\hat{X}_1,\hat{X}_2),I(X;\hat{X}_1\|\hat{X}_2)+R_{12}),\{I(X;\hat{X}_2)-R_{12}\}^+)$

TABLE VIII
THE CORNER POINTS OF THE ACHIEVABILITIES OF THE MAC WITH ONE COMMON MESSAGE AND ONE PRIVATE MESSAGE AND CONFERENCING ENCODERS AND OF SUCCESSIVE REFINEMENT WITH CONFERENCING DECODERS.

VI. CONCLUSION

In this paper, we introduced new models of cooperation in multi terminal source coding. The setting of successive refinement with single encoder and two decoders was generalized to incorporate cooperation between the users via (a) *conferencing*, or (b) *cribbing*. A new scheme,“Forward Encoding” and “Block Markov Decoding” was used to derive the rate regions for strictly-causal and causal cribbing. Certain numerical examples are presented and show how cooperation via cribbing can boost the rate region. Finally, we introduce dual channel coding problems, and establish duality between successive refinement with cribbing decoders and communication over the MAC with common message and cribbing encoders.

REFERENCES

[1] F. Willems and E. van der Meulen, “The discrete memoryless multiple-access channel with cribbing encoders,” *Information Theory, IEEE Transactions on*, vol. 31, no. 3, pp. 313 – 327, May 1985.

[2] H. H. Permuter and H. Asnani, “Multiple access channel with partial and controlled cribbing encoders,” *CoRR*, vol. abs/1103.4007, 2011.

[3] F. Willems, “The discrete memoryless multiple access channel with partially cooperating encoders (corresp.),” *Information Theory, IEEE Transactions on*, vol. 29, no. 3, pp. 441 – 445, May 1983.

[4] S. Bross, A. Lapidoth, and M. Wigger, “The gaussian mac with conferencing encoders,” in *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, July 2008, pp. 2702 –2706.

[5] M. Wiese, H. Boche, I. Bjelakovic, and V. Jungnickel, “The compound multiple access channel with partially cooperating encoders,” *Information Theory, IEEE Transactions on*, vol. 57, no. 5, pp. 3045 –3066, May 2011.

[6] O. Simeone, D. Gunduz, H. Poor, A. Goldsmith, and S. Shamai, “Compound multiple-access channels with partial cooperation,” *Information Theory, IEEE Transactions on*, vol. 55, no. 6, pp. 2425 –2441, June 2009.

[7] I. Maric, R. Yates, and G. Kramer, “The discrete memoryless compound multiple access channel with conferencing encoders,” in *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, Sept. 2005, pp. 407 –410.

[8] S. I. Bross, Y. Steinberg, and S. Tinguely, "The causal cognitive interference channel," *International Zurich Seminar on Communciations (IZS)*, March 2010.

[9] V. Prabhakaran and P. Viswanath, "Interference channels with source cooperation," *Information Theory, IEEE Transactions on*, vol. 57, no. 1, pp. 156 –186, Jan. 2011.

[10] ——, "Interference channels with destination cooperation," *Information Theory, IEEE Transactions on*, vol. 57, no. 1, pp. 187 –209, Jan. 2011.

[11] C. T. K. Ng and A. J. Goldsmith, "Capacity and cooperation in wireless networks," *Information Theory and Applications (ITA)*, 2006.

[12] H. Yamamoto, "Source coding theory for cascade and branching communication systems," *Information Theory, IEEE Transactions on*, vol. 27, no. 3, pp. 299 – 308, May 1981.

[13] P. Cuff, H.-I. Su, and A. El Gamal, "Cascade multiterminal source coding," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, 28 2009-July 3 2009, pp. 1199 –1203.

[14] H. H. Permuter and T. Weissman, "Cascade and triangular source coding with side information at the first two nodes," *CoRR*, vol. abs/1001.1679, 2010.

[15] Y.-K. Chia, H. H. Permuter, and T. Weissman, "Cascade, triangular and two way source coding with degraded side information at the second user," *CoRR*, vol. abs/1010.3726, 2010.

[16] H. Yamamoto, "Source coding theory for a triangular communication system," *Information Theory, IEEE Transactions on*, vol. 42, no. 3, pp. 848 –853, May 1996.

[17] P. Cuff and L. Zhao, "Coordination using implicit communication," *CoRR*, vol. abs/1108.3652, 2011.

[18] W. Equitz and T. Cover, "Successive refinement of information," *Information Theory, IEEE Transactions on*, vol. 37, no. 2, pp. 269 –275, Mar 1991.

[19] A. Aaron, D. Varodayan, and B. Girod, "Wyner-ziv residual coding of video," in *Proc. Picture Coding Symposium, PCS-2006*, 2006.

[20] D. Vasudevan, C. Tian, and S. N. Diggavi, "Lossy source coding for a cascade communication system with side-informations," *Proceedings of Allerton Conference on Communication, Control, and Computing*, 2006.

[21] A. E. Gamal and Y.-H. Kim, "Lecture notes on network information theory," *CoRR*, vol. abs/1001.3404, 2010.

[22] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *Information and Decision Processes*, R. E. Machol, Ed. McGraw-Hill, 1960, pp. 93–126.

[23] S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1181 – 1203, May 2003.

[24] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1629–1638, June 2002.

[25] W. Yu, "Duality and the value of cooperation in distributive source and channel coding problems," in *Communication, Control, and Computing, 2003. Allerton 2003. 41th Annual Allerton Conference on*, 2003.

[26] A. Shirazi, U. Basher, and H. Permuter, "Channel coding and source coding with increased partial side information," in

Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on, 29 2010-Oct. 1 2010, pp. 753 –759.

[27] K. Kittichokechai, T. J. Oechtering, M. Skoglund, and R. Thobaben, “Source and channel coding with action-dependent partially known two-sided state information,” in *Proceedings IEEE International Symposium on Information Theory*, Jun. 2010, pp. 629–633.

[28] A. Gupta and S. S. Verdú and, “Operational duality between gelfand-pinsker and wyner-ziv coding,” in *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2010, pp. 530 –534.

[29] P. Cuff, H. H. Permuter, and T. M. Cover, “Coordination capacity,” *IEEE Trans. Inf. Theory*, vol. 59, pp. 4181–4206, Sep. 2010.

[30] D. Slepian and J. K. Wolf, “A coding theorem for multiple-access channel with correlated sources,” *Bell Syst. Tech. J.*, vol. 51, p. 10371076, 1973.

[31] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New-York: Wiley, 2006.

[32] F. M. J. Willems, “The discrete memoryless multiple channel with partially cooperating encoders,” *IEEE Trans. Inf. Theory*, vol. 29, no. 6, p. 441445, 1983.

[33] N. Lauritzen, “Lectures on convex sets,” 2010, availble at http://home.imf.au.dk/niels/lecconset.pdf.

APPENDIX A
SUCCESSIVE REFINEMENT WITH CONFERENCING DECODERS, FIG. 3

Consider Fig. 3, here Decoder 1 cooperates with Decoder 2 by providing an additional description T_{12} to it. The rate region is given by,

$$R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2) \tag{118}$$

$$R_0 + R_{12} \geq I(X; \hat{X}_2), \tag{119}$$

for some joint probability distribution $P_{X,\hat{X}_1,\hat{X}_2}$ such that $\mathsf{E}[d_i(X_i, \hat{X}_i)] \leq D_i$, for $i = 1, 2$. We will briefly describe the proof as they are based on standard arguments used throughout the paper.

Achievability : We provide the achievability under two cases,

- *Case 1* : $R_1 \leq R_{12}$, here we describe T_1 through T_{12}, thus both the decoders know (T_0, T_1) and hence the following region is achievable,

$$R_1 \leq R_{12} \tag{120}$$

$$R_0 + R_1 \geq I(X; \hat{X}_1, \hat{X}_2), \tag{121}$$

for a joint probability distribution $P_{X,\hat{X}_1,\hat{X}_2}$ such that distortion constraints are satisfied. The region is equivalent to, (call it Region 1)

$$R_1 \leq R_{12} \tag{122}$$

$$R_0 + R_1 \geq I(X;\hat{X}_1,\hat{X}_2) \tag{123}$$

$$R_0 + R_{12} \geq I(X;\hat{X}_2). \tag{124}$$

- *Case 2* : $R_1 > R_{12}$, here T_1 is described as a tuple (T_1', T_1'') of rate $(R_1 - R_{12}, R_{12})$, and T_1'' is described via the conferencing link. Thus this problem is similar to original successive refinement problem, where encoder has a private rate $R_1 - R_{12}$ and a common rate $R_0 + R_{12}$, and hence the following region (call Region 2) is achievable (follows from the achievability of Equitz and Cover [18]),

$$R_1 > R_{12} \tag{125}$$

$$R_0 + R_1 \geq I(X;\hat{X}_1,\hat{X}_2) \tag{126}$$

$$R_0 + R_{12} \geq I(X;\hat{X}_2), \tag{127}$$

for a joint probability distribution $P_{X,\hat{X}_1,\hat{X}_2}$ such that distortion constraints are satisfied.

We finish the proof of achievability by combining Region 1 and Region 2. Converse follows from standard cutset bound arguments and is omitted.

APPENDIX B
PROOF OF LEMMA 1

The proof is done by proving set inclusions in two directions as done for Theorem 3 in [20]. First we prove, $\tilde{\mathcal{R}}(D_1, D_2) \subseteq \mathcal{R}_{cascade}(D_1, D_2)$. Suppose a pair $(\tilde{R}_0, \tilde{R}_0 + \tilde{R}_1) \in \tilde{\mathcal{R}}(D_1, D_2)$. This implies there exists a $(2^{n\tilde{R}_0}, 2^{n\tilde{R}_1}, n)$ code (cf. Definition 2), for the setting of successive refinement with cribbing (Fig. 4), such that distortion constraints $D_1 + \epsilon$ and $D_2 + \epsilon$ are met at the decoders. We can use this code to generate a code of rates $R_1 = \tilde{R}_0 + \tilde{R}_1$ and $R_{12} = \tilde{R}_0$, for our cascade setting with cribbing (Fig. 5), with exactly same distortions at the decoders. This proves one direction.

For the other direction, i.e., $\mathcal{R}_{cascade}(D_1, D_2) \subseteq \tilde{\mathcal{R}}(D_1, D_2)$, assume, $(R_{12}, R_1) \in \mathcal{R}_{cascade}(D_1, D_2)$,

which means there exist codes with rates (R_{12}, R_1) with decoders incurring distortions, $D_1+\epsilon$ and $D_2+\epsilon$. Assume the messages sent on first and second link in our cascade problem be T_1 and T_{12} respectively. T_{12} is a function of T_1, and we have

$$nR_{12} \geq H(T_{12}) \tag{128}$$

$$nR_1 \geq H(T_1) = H(T_1, T_{12}) = H(T_{12}) + H(T_1|T_{12}). \tag{129}$$

Using this code, we now will construct a code for successive refinement setting with cribbing decoders. Specifically, we consider encoding in B blocks where each block is of length n. Denote by $T_{12}(i)$ and $T_1(i)$ the messages which are transmitted in the cascade source coding setting in i^{th} block. Note that the tuple $\{T_{12}(1), \cdots, T_{12}(B)\}$ can be communicated to both decoders in the successive refinement setting with vanishing probability of error, with a rate $R_0 = \frac{1}{n}H(T_{12})$, with large number of blocks, and similarly the tuple $(T_1(1), \cdots, T_1(B))$ can be communicated to Decoder 1 with rate (using Slepian Wolf Coding as $(T_{12}(i), T_1(i))$ are independent) $R_1 = \frac{1}{n}H(T_1|T_{12})$. Thus Decoder 1 and Decoder 2 will know exactly the same (T_{12}, T_1) and T_{12} respectively as they would know in cascade setting. Since the cribbing structure (Decoder 2 gets the crib from Decoder 1 non-causally, strictly-causally and causally) is same in cascade source coding and successive refinement setting, decoders will be able to achieve same distortion levels, (D_1, D_2). This implies, $(\frac{1}{n}H(T_{12}), \frac{1}{n}H(T_1|T_{12})) \in \mathcal{R}(D_1, D_2)$ or $(\frac{1}{n}H(T_{12}), \frac{1}{n}H(T_1)) \in \tilde{\mathcal{R}}(D_1, D_2)$, which implies by Eq. (128)-(129), that $(R_0, R_1) \in \tilde{\mathcal{R}}(D_1, D_2)$.

APPENDIX C
PROOF OF ACHIEVABILITY IN THEOREM 2

We describe in detail the achievablility in Theorem 2.

- *Codebook Generation* : Fix the distribution $P_X P_{\hat{Z}_1, \hat{X}_2|X} P_{\hat{X}_1|X, \hat{Z}_1, \hat{X}_2}$, $\epsilon > 0$ such that $E[d_1(X, \hat{X}_1)] \leq \frac{D_1}{1+\epsilon}$ and $E[d_2(X, \hat{X}_2)] \leq \frac{D_2}{1+\epsilon}$. Generate codebook $\mathcal{C}_{\hat{X}_2}$ consisting of $2^{nI(X;\hat{X}_2)}$ $\hat{X}_2^n(m_h)$ codewords generated i.i.d $\sim P_{\hat{X}_2}$, $m_h \in [1 : 2^{nI(X;\hat{X}_2)}]$. For each m_h, generate a codebook $\mathcal{C}_{\hat{Z}_1}(m_h)$ consisting of $2^{nI(X;\hat{Z}_1|\hat{X}_2)}$ $\hat{Z}_1^n$ codewords generated i.i.d. $\sim P_{\hat{Z}_1|\hat{X}_2}$. We then bin these generated $\hat{Z}_1^n$ codewords for each m_h, in 2^{nR_0} vertical bins, $\mathcal{B}(m_v)$, $m_v \in [1 : 2^{nR_0}]$ and index them accordingly with $l \in [1 : 2^{n(I(X;\hat{Z}_1|\hat{X}_2)-R_0)}]$. $\hat{Z}_1^n$ codewords can be indexed equivalently as the tuple (m_h, m_v, l). For

each $\hat{Z}_1^n(m_h, m_v, l)$ codeword, generate a codebook, $\mathcal{C}_{\hat{X}_1}(m_h, m_v, l)$ consisting of $2^{nI(X;\hat{X}_1|\hat{Z}_1,\hat{X}_2)}$ $\hat{X}_1^n(m_h, m_v, l, k)$ codewords generated i.i.d. $\sim P_{\hat{X}_1|\hat{Z}_1,\hat{X}_2}$, $k \in [1 : 2^{nI(X;\hat{X}_1|\hat{Z}_1,\hat{X}_2)}]$. Thus the generation of codebooks is similar to that in perfect cribbing, except here we generate one more layer, of $\hat{Z}_1$ codewords. Also we bin $\hat{Z}_1^n$ codewords instead of $\hat{X}_1^n$. Here, m_h and m_v correspond to the row and column index of the "doubly-indexed" bin which contains $\hat{Z}_1^n$ codeword and for each $\hat{Z}_1^n$ codeword, a codebook of $\hat{X}_1^n$ codebook is generated.

- *Encoding* : Given source sequence X^n, encoder finds the index $m_h \in [1 : 2^{nI(X;\hat{X}_2)}]$ from codebook $\mathcal{C}_{\hat{X}_2}$ such that $(X^n, \hat{X}_2^n(m_h)) \in \mathcal{T}_\epsilon^n$. The encoder then finds the index tuple (m_v, l) from the $\mathcal{C}_{\hat{Z}_1}(m_h)$ codebook, such that $(X^n, \hat{Z}_1^n(m_h, m_v, l), \hat{X}_2^n(m_h)) \in \mathcal{T}_\epsilon^n$. Encoder then finds the index k from the $\mathcal{C}_{\hat{X}_1}(m_h, m_v, l)$ codebook, such that $(X^n, \hat{X}_1^n(m_h, m_v, l, k), \hat{Z}_1^n(m_h, m_v, l), \hat{X}_2^n(m_h)) \in \mathcal{T}_\epsilon^n$. Thus $\hat{Z}_1^n \in \mathcal{B}(m_v)$. m_v is described as R_0 and the index triple, (m_h, l, k) is described as R_1, thus

$$\begin{aligned} R_1 &\geq I(X;\hat{X}_2) + I(X;\hat{Z}_1|\hat{X}_2) - R_0 + I(X;\hat{X}_1|\hat{Z}_1,\hat{X}_2) \\ \text{or,} \quad R_0 + R_1 &\geq I(X;\hat{Z}_1,\hat{X}_1,\hat{X}_2) = I(X;\hat{X}_1,\hat{X}_2), \end{aligned} \tag{130}$$

as $\hat{Z}_1 = g(\hat{X}_1)$.

- *Decoding* : Using the indices sent by encoder, Decoder 1 constructs $\hat{X}_1^n = \hat{X}_1^n(m_h, m_v, l, k)$. Decoder 2 gets $\hat{Z}_1^n$ and column index m_v, and infers the unique index m_h such that $\hat{Z}_1^n = \hat{Z}_1^n(m_h, m_v, \tilde{l})$ for some $\tilde{l} \in [1 : 2^{n(I(X;\hat{Z}_1|\hat{X}_2)-R_0)}]$.
- *Distortion Analysis* : Consider the following events :

  –

  $$\mathcal{E}_0 = \text{Encoder cannot find } (\hat{X}_2^n, \hat{Z}_1^n, \hat{X}_1^n) \text{ jointly typical with given source } X^n \tag{131}$$

  But the probability of this event vanishes by *Covering Lemma*, Lemma 3 as there are $2^{nI(X;\hat{X}_2)}$ $\hat{X}_2^n$ codewords, for each $\hat{X}_2^n$ codeword there are $2^{nI(X;\hat{Z}_1|\hat{X}_2)}$ $\hat{Z}_1^n$ codewords and finally for each $\hat{Z}_1^n$ codeword there are $2^{nI(X;\hat{X}_1|\hat{Z}_1,\hat{X}_2)}$ $\hat{X}_1^n$ codewords . Without loss of generality, now suppose that $(m_h, m_v, l, k) = (1, 1, 1, 1)$ was sent by the encoder.

–

$$\mathcal{E}_1 = \hat{Z}_1^n \text{ does not lie in bin with row index } m_h = 1 \text{ and column index } m_v = 1 \quad (132)$$

$$= \left\{ \hat{Z}_1^n \neq \hat{Z}_1^n(1,1,\tilde{l}), \text{ for any } \tilde{l} \in [1 : 2^{n(I(X;\hat{Z}_1|\hat{X}_2)-R_0)}] \right\}. \quad (133)$$

But the probability of this event goes to zero, because of our encoding procedure, as $\hat{Z}_1^n = \hat{Z}_1^n(1,1,1)$.

–

$$\mathcal{E}_2 = \hat{Z}_1^n \text{ lies in bin with row index } \hat{m}_h \neq 1 \text{ and column index } m_v = 1. \quad (134)$$

$$= \left\{ \hat{Z}_1^n = \hat{Z}_1^n(\hat{m}_h,1,\tilde{l}), \hat{m}_h \neq 1, \text{ for some } \tilde{l} \in [1 : 2^{n(I(X;\hat{Z}_1|\hat{X}_2)-R_0)}] \right\}. \quad (135)$$

Using similar argument as in the case of perfect cribbing, probability of this event goes to zero with large n, if

$$I(X;\hat{X}_1,\hat{X}_2) - R_0 \leq I(\hat{Z}_1;\hat{Z}_1,\hat{X}_2) = H(\hat{Z}_1). \quad (136)$$

Thus consider the event, $\mathcal{E} = \mathcal{E}_0 \cup \mathcal{E}_1 \cup \mathcal{E}_2$, using Eq. (130) and Eq. (136), probability of this event goes to zero with large n if,

$$R_0 + R_1 \geq I(X;\hat{X}_1,\hat{X}_2) \quad (137)$$

$$R_0 \geq \{I(X;\hat{Z}_1,\hat{X}_2) - H(\hat{Z}_1)\}^+. \quad (138)$$

Distortion is bounded as in other sections.

APPENDIX D

PROOF OF THEOREM 7, MAC WITH CRIBBING ENCODERS AND COMMON MESSAGE

Proof of achievability of Theorem 7, noncausal case: The main idea of the achievability proof is to split message m_1 into two parts m_1' and m_1'' with rates R_1' and R_1'' respectively, such that $R_1 = R_1' + R_1''$. Message m_1' is transmitted to Encoder 1 through the cribbing signal Z_1^n, while m_1'' remains as a private message to Encoder 1.

Code design: For the given joint distribution $P(x_1, x_2)$ generate $2^{nR_1'}$ codewords z_1^n distributed i.i.d. according to $P(z_1)$. For each codeword z_1^n generate 2^{nR_0} codewords x_2^n according to $P(x_2|z_1)$. For each codewords pair (z_1^n, x_2^n) generate $2^{nR_1''}$ x_1^n codewords according to $P(x_1|z_1, x_2)$.

Encoding and decoding:

- Encoder 1: maps (m_1', m_1'', m_0) to $(z_1^n(m_1'), x_2^n(z_1^n, m_0), x_1^n(x_2^n, z_1^n, m_1''))$, and transmits $x_1^n(x_2^n, z_1^n, m_1'')$.
- Encoder 2: transmits $x_2^n(z_1^n, m_0)$.
- Decoder: looks for $(\hat{m}_0, \hat{m}_1', \hat{m}_1'')$ such that

$$(z_1^n(\hat{m}_1'), x_2^n(z_1^n, \hat{m}_0), x_1^n(x_2^n, z_1^n, \hat{m}_1''), y^n) \in T_\epsilon^{(n)}. \tag{139}$$

Error analysis: Without loss of generality let's assume that the message that is sent is $m_0 = 1, m_1' = 1$, and $m_1'' = 1$.

- Let E_0 be the event that $(x_1^n(1), x_2^n(1), y^n) \notin T_\epsilon^{(n)}$. Clearly, $\Pr\{E_0\} \to 0$ by the law of large numbers. Hence, for the rest of the events we can assume that $(x_1^n(1), x_2^n(1)) \in T_\epsilon^{(n)}$.
- Let $E_{1,j}$ be the event that $z_1^n(1) = z_1^n(j)$. And let E_1 be the event that there exists an $j \neq 1$ such that $z_1^n(1) = z_1^n(j)$. Following from the definition $E_1 = \cup_{j\geq 1} E_{1,j}$. Let's bound the probability of E_1 using the union bound and the fact that $\Pr\{E_{b,j}\} \leq 2^{-n(H(Z_1)-\epsilon)}$.

$$\begin{aligned}
\Pr\{E_1\} &= \Pr\{\cup_{j\geq 1} E_{1,j}\} && (140)\\
&\leq \sum_{i\geq 2} \Pr\{E_{1,j}\} && (141)\\
&\leq \sum_{i\geq 2} 2^{-n(H(Z_1)-\epsilon)} && (142)\\
&= 2^{n(R_1' - H(Z_1)+\epsilon)}, && (143)
\end{aligned}$$

hence, if

$$R_1' < H(Z_1), \tag{144}$$

$\Pr\{E_1\} \to 0$ as $n \to \infty$.

- Let $E_{i,j,k}$ be the event probability that for $\hat{m}_1' = i, \hat{m}_0 = j$, and $\hat{m}_1'' = k$

$$(z_1^n(\hat{m}_1'), x_2^n(z_1^n, \hat{m}_0), x_1^n(x_2^n, z_1^n, \hat{m}_1''), y^n) \in T_\epsilon^{(n)}. \tag{145}$$

Let E_3 be the event that exists an $(i,j,k) \neq (1,1,1)$ such that $E_{i,j,k}$ occurs.

$$\Pr\{E_3\} \leq \Pr\{\bigcup_{i\geq 2, j\geq 1, k\geq 1} E_{i,j,k}\} + \Pr\{\bigcup_{i=1, j\geq 2, k\geq 1} E_{i,j,k}\} + \Pr\{\bigcup_{i=1, j=1, k\geq 2} E_{i,j,k}\}. \tag{146}$$

Now let's bound each term. Consider the first term in the RHS of (146)

$$\begin{aligned} \Pr\{\bigcup_{i\geq 2, j\geq 1, k\geq 1} E_{i,j,k}\} &\leq \sum_{i=2,j=1,k=1}^{2^{nR_1'}, 2^{nR_0}, 2^{nR_1''}} 2^{-n(I(Z_1,X_1,X_2;Y)-\epsilon)} \\ &\leq 2^{n(R_0+R_1'+R_1''-I(Z_1,X_1,X_2;Y)+\epsilon)}, \end{aligned} \tag{147}$$

hence if

$$R_0 + R_1 < I(X_1, X_2; Y), \tag{148}$$

then the probability above goes to zero. Consider the second term in the RHS of (146)

$$\begin{aligned} \Pr\{\bigcup_{i=1, j\geq 2, k\geq 1} E_{i,j,k}\} &\leq \sum_{j=2,k=1}^{2^{nR_0}, 2^{nR_1''}} 2^{-n(I(X_1,X_2;Y|Z_1)-\epsilon)} \\ &\leq 2^{n(R_0+R_1''-I(X_1,X_2;Y|Z_1)+\epsilon)}, \end{aligned} \tag{149}$$

hence if

$$R_0 + R_1'' < I(X_1, X_2; Y|Z_1), \tag{150}$$

then the probability above goes to zero.

Consider the third term in the RHS of (146)

$$\Pr\{\bigcup_{i=1, j=1, k\geq 2} E_{i,j,k}\} \leq 2^{n(R_1''-I(X_1;Y|Z_1,X_2)+\epsilon)}, \tag{151}$$

hence if

$$R_1'' < I(X_1; Y|Z_1, X_2), \tag{152}$$

then the probability above goes to zero.

Gathering (144), (148), (150) and (152) we obtain

$$\begin{aligned} R_1' &< H(Z_1) && (153)\\ R_0+R_1 &< I(X_1,X_2;Y) && (154)\\ R_0+R_1'' &< I(X_1,X_2;Y|Z_1) && (155)\\ R_1'' &< I(X_1;Y|Z_1,X_2). && (156) \end{aligned}$$

Using Fourier−Motzkin elimination [33] we obtain

$$\begin{aligned} R_0+R_1 &< I(X_1,X_2;Y) && (157)\\ R_0+R_1 &< I(X_1,X_2;Y|Z_1)+H(Z_1) && (158)\\ R_1 &< I(X_1;Y|Z_1,X_2)+H(Z_1). && (159) \end{aligned}$$

Since $I(X_1,X_2;Y) \leq I(X_1,X_2;Y|Z_1)+H(Z_1)$ the second inequality in (187) is redundant and therefore the region

$$\begin{aligned} R_0+R_1 &< I(X_1,X_2;Y)\\ R_1 &< I(X_1;Y|Z_1,X_2)+H(Z_1). && (160) \end{aligned}$$

is achievable. ■

Proof of converse for the non causal case: Let $(2^{nR_0},2^{nR_1},n)$ be a non causal cribbing MAC code as defined in Def. 4 with a probability of error $P_e^{(n)}$. Consider,

$$\begin{aligned} R_0+R_1 &= H(M_0,M_1) && (161)\\ &= I(M_0,M_1;Y^n)+H(M_0,M_1|Y^n) && (162)\\ &\overset{(a)}{\leq} I(X_1^n,X_2^n;Y^n)+n\epsilon_n && (163)\\ &\overset{(b)}{\leq} \sum_{i=1}^{n} I(X_{1,i},X_{2,i};Y_i)+n\epsilon_n && (164)\\ &\overset{(c)}{\leq} nI(X_{1,Q},X_{2,Q};Y_Q|Q)+n\epsilon_n && (165)\\ &\leq nI(X_{1,Q},X_{2,Q};Y_Q)+n\epsilon_n, && (166) \end{aligned}$$

where (a) follows from Fano's inequality where $\epsilon_n = (\frac{1}{n} + R_0 + R_1)P_e^{(n)}$, step (b) follows from the memoryless nature of the MAC and (c) follows from denoting Q as uniform random variable over the alphabet $\{1, 2, ..., n\}$. Now consider

$$
\begin{aligned}
R_1 &= H(M_1) && (167)\\
&= H(M_1|M_0) && (168)\\
&\stackrel{(a)}{\leq} I(M_1; Y^n|M_0) + n\epsilon_n && (169)\\
&\leq I(X_1^n, Z_1^n; Y^n|X_2^n) + n\epsilon_n && (170)\\
&= I(Z_1^n; Y^n|X_2^n) + I(X_1^n; Y^n|X_2^n, Z_1^n) + n\epsilon_n && (171)\\
&\stackrel{(b)}{\leq} \sum_{i=1}^{n} H(Z_{1,i}) + I(X_{1,i}; Y_i|X_{2,i}, Z_{1,i}) + n\epsilon_n && (172)\\
&\stackrel{(c)}{\leq} \sum_{i=1}^{n} H(Z_{1,Q}|Q) + I(X_{1,Q}; Y_Q|X_{2,Q}, Z_{1,Q}, Q) + n\epsilon_n && (173)\\
&\leq \sum_{i=1}^{n} H(Z_{1,Q}) + I(X_{1,Q}; Y_Q|X_{2,Q}, Z_{1,Q}) + n\epsilon_n, && (174)
\end{aligned}
$$

where the justification for (a), (b) and (c) follows from similar arguments as steps (a), (b) and (c) for bounding $R_0 + R_1$. Since the rate pair is achievable, the code type is arbitrary close to the restricted distribution $P(x_1, x_2)$ and using Lemma 5 we conclude that the distribution of $X_{1,Q}$, $X_{2,Q}$ is arbitrary close to the restricted distribution $P(x_1, x_2)$. Finally, by denoting $Z_1 = Z_Q$, $X_1 = X_{1,Q}$, $X_2 = X_{2,Q}$ and $Y = Y_Q$ and taking into account that $P_e^{(n)}$ is going to zero as $n \to \infty$ we obtain that the region $\mathcal{R}^{nc}(P)$ upper bound the capacity region. ■

Proof of achievability of Theorem 7, strictly causal case: The main idea of the achievability proof is to combine the rate splitting idea that we used in the noncausal case with the Markov block coding. We assume that the transmission is done in a block of size nB where B is the number of subblocks and each subblock is of length n. Let $m_{0,b}, m_{1,b}$ be the messages sent in block b. Similarly to the noncausal case, split message $m_{1,b}$ into two parts $m'_{1,b}$ and $m''_{1,b}$ with rates R'_1 and R''_1 respectively, such that $R_1 = R'_1 + R''_1$. Message $m'_{1,b}$ is transmitted to Encoder 1 through the cribbing signal, while $m''_{1,b}$ remains as a private message to Encoder 1. Because of the causality, the the message $m'_{1,b}$ is known to Encoder 2 only at the end of block b.

Code design: For fixed a joint distribution $P(x_1, x_2)$ generate $2^{n(R_0+R_1')}$ codewords x_2^n each associated with the pair of messages $(m_{0,b}, m'_{1,b-1})$. For each codeword x_2^n generate $2^{nR_1'}$ codewords z_1^n according to conditional distribution $P(z_1|x_2)$ associated with $m'_{1,b}$. For each codeword pair (z_1^n, x_2^n) generate $2^{nR_1''}$ codewords x_1^n according to conditional distribution $P(x_1|z_1, x_2)$ associated with $m''_{1,b}$.

Encoding and decoding:

- Encoder 1: In block b maps $(m'_{1,b-1}, m'_{1,b}, m''_{1,b}, m_{0,b})$ to $(x_2^n(m_{0,b}, m'_{1,b-1}), z_1^n(m'_{1,b}, x_2^n), x_1^n(m''_{1,b}, x_2^n, z_1^n))$, and transmits $x_1^n(m''_{1,b}, x_2^n, z_1^n))$.
- Encoder 2: Transmits $x_2^n(m_{0,b}, m'_{1,b-1})$. Message $m'_{1,b-1}$ is known to Encoder 2 since at the end of block $b-1$, $z_1^n(m'_{1,b-1}, x_2^n)$ and x_2^n are known.
- Decoder: Does backward decoding. We assume that when decoding block b message $m'_{1,b}$ is known and it looks for tuple $(\hat{m}_{0,b}, \hat{m}'_{1,b-1}, \hat{m}''_{1,b})$ such that

$$(x_2^n(\hat{m}_{0,b}, \hat{m}'_{1,b-1}), z_1^n(m'_{1,b}, x_2^n), x_1^n(\hat{m}''_{1,b}, x_2^n, z_1^n), y^n) \in T_\epsilon^{(n)}. \tag{175}$$

Error analysis: Without loss of generality let's assume that the message that is sent is $m_{0,b} = 1, m'_{1,b} = 1, m'_{1,b-1} = 1$, and $m''_{1,b} = 1$.

- Let E_0 be the event that $(x_1^n(1), x_2^n(1)) \notin T_\epsilon^{(n)}$. Clearly, $\Pr\{E_0\} \to 0$ by the law of large numbers. Hence, for the rest of the events we can assume that $(x_1^n(1), x_2^n(1)) \in T_\epsilon^{(n)}$.
- Let E_1 be the event that in block $b-1$ there exists an $j \neq 1$, such that $z_1^n(1) = z_1^n(j)$ for some codeword x_2^n. Similar to the analysis for the noncausal case

$$\Pr\{E_1\} = 2^{n(R_1' - H(Z_1|X_2)+\epsilon)}, \tag{176}$$

hence, if

$$R_1' < H(Z_1|X_2), \tag{177}$$

$\Pr\{E_1\} \to 0$ as $n \to \infty$.
- Let $E_{i,j,k}$ be the event probability that for $\hat{m}'_{1,b-1} = i, \hat{m}_{0,b} = j$, and $\hat{m}''_{1,b} = k$, given that $m'_{1,b}$ is

known correctly from pervious subblock decoding:

$$(x_2^n(\hat{m}_{0,b},\hat{m}'_{1,b-1}),z_1^n(m'_{1,b},x_2^n),x_1^n(\hat{m}''_{1,b},x_2^n,z_1^n),y^n)\in T_\epsilon^{(n)}. \tag{178}$$

Let E_3 be the event that exists an $(i,j,k)\neq(1,1,1)$ such that $E_{i,j,k}$ occurs.

$$\Pr\{E_3\}\leq\Pr\{\bigcup_{(i,j)\neq(1,1),k\geq 1}E_{i,j,k}\}+\Pr\{\bigcup_{(i,j)=(1,1),k\geq 2}E_{i,j,k}\}. \tag{179}$$

Now let's bound each term. Consider the first term in the RHS of (179)

$$\Pr\{\bigcup_{(i,j)\neq(1,1),k\geq 1}E_{i,j,k}\} \leq 2^{n(R_0+R_1-I(Z_1,X_1,X_2;Y)+\epsilon)}, \tag{180}$$

hence if

$$R_0+R_1<I(X_1,X_2;Y) \tag{181}$$

then the probability above goes to zero. Consider the second term in the RHS of (179)

$$\Pr\{\bigcup_{(i,j)=(1,1),k\geq 2}E_{i,j,k}\} \leq 2^{n(R''_1-I(X_1;Y|Z_1,X_2)+\epsilon)}, \tag{182}$$

hence if

$$R''_1<I(X_1;Y|Z_1,X_2), \tag{183}$$

then the probability above goes to zero.

Gathering (177), (181), and (183) we obtain

$$\begin{aligned} R'_1 &< H(Z_1|X_2) && (184)\\ R_0+R_1 &< I(X_1,X_2;Y) && (185)\\ R''_1 &< I(X_1;Y|Z_1,X_2). && (186)\end{aligned}$$

Using Fourier$-$Motzkin elimination

$$\begin{aligned} R_0+R_1 &< I(X_1,X_2;Y) && (187)\\ R_1 &< I(X_1;Y|Z_1,X_2)+H(Z_1|X_2). && (188)\end{aligned}$$

is achievable. ■

Proof of converse for the strictly causal case: Let $(2^{nR_1}, 2^{nR_0}, n)$ be a strictly causal cribbing MAC code as defined in Def. 4 with a probability of error $P_e^{(n)}$. Following the exact same steps as in the converse of the noncausal case in (189) we obtain

$$R_0 + R_1 \quad \leq \quad nI(X_{1,Q}, X_{2,Q}; Y_Q) + n\epsilon_n. \tag{189}$$

Following the exact same first four steps as in converse of the non causal case to bound R_1, (167) we obtain

$$\begin{aligned} R_1 \quad &\leq \quad I(Z_1^n; Y^n | X_2^n) + I(X_1^n; Y^n | X_2^n, Z_1^n) + n\epsilon_n & (190)\\ &\leq \quad \sum_{i=1}^{n} H(Z_{1,i} | X_{2,i}) + I(X_{1,i}; Y_i | X_{2,i}, Z_{1,i}) + n\epsilon_n & (191)\\ &\leq \quad \sum_{i=1}^{n} H(Z_{1,Q} | X_{2,Q}) + I(X_{1,Q}; Y_Q | X_{2,Q}, Z_{1,Q}) + n\epsilon_n, & (192) \end{aligned}$$

Since the rate pair is achievable, the code type is arbitrary close to the restricted distribution $P(x_1, x_2)$ and using Lemma 5 we conclude that the distribution of $X_{1,Q}$, $X_{2,Q}$ is arbitrary close to the restricted distribution $P(x_1, x_2)$. Finally, by denoting $Z_1 = Z_Q$, $X_1 = X_{1,Q}$, $X_2 = X_{2,Q}$ and $Y = Y_Q$ and taking into account that $P_e^{(n)}$ is going to zero as $n \to \infty$ we obtain that the region $\mathcal{R}^{nc}(P)$ upper bound the capacity region. ■

Proof of achievability of Theorem 7, causal case: In this proof we show how the causal case achievability follows directly from the proof of the strictly causal case with one modification: instead of codewords x_2^n we generate codewords u^n, and the input to the channel is $x_{2,i} = f(u_i, x_{1,i})$. This is possible since Encoder 2 observes causally the signal from Encoder 1. By replacing X_2 with U in $\mathcal{R}^{sc}(P)$ and applying $x_{2,i} = f(u_i, z_{1,i})$ and taking into account the equality $I(Y; X_1, U) = I((Y; X_1, U, f(Z_1, U)) = I(Y; X_1, X_2)$ we obtain the region $\mathcal{R}^{c}(P)$. ■

Proof of converse for the causal case: Let $(2^{nR_0}, 2^{nR_1}, n)$ be a partial strictly causal cribbing MAC code as defined in Def. 4 with a probability of error $P_e^{(n)}$. Following the exact same steps as in the converse of the noncausal case in (189) we obtain

$$R_0 + R_1 \quad \leq \quad nI(X_{1,Q}, X_{2,Q}; Y_Q) + n\epsilon_n, \tag{193}$$

Now consider

$$
\begin{aligned}
R_1 &= H(M_1) && (194)\\
&= H(M_1|M_0) && (195)\\
&\overset{(a)}{\leq} I(M_1;Y^n|M_0)+n\epsilon_n && (196)\\
&\leq I(X_1^n,Z_1^n;Y^n|M_0)+n\epsilon_n && (197)\\
&= I(Z_1^n;Y^n|M_0)+I(X_1^n;Y^n|M_0,Z_1^n)+n\epsilon_n && (198)\\
&\overset{(b)}{\leq} \sum_{i=1}^{n} I(Z_{1,i};Y^n|M_0,Z^{i-1})+I(X_{1,i};Y_i|X_{2,i},M_0,Z_1^n,X^{i-1})+n\epsilon_n && (199)\\
&\overset{(c)}{\leq} \sum_{i=1}^{n} H(Z_{1,i}|M_0,Z^{i-1})+I(X_{1,i};Y_i|X_{2,i},M_0,Z_1^{i-1})+n\epsilon_n && (200)\\
&\leq \sum_{i=1}^{n} H(Z_{1,Q}|U_Q)+I(X_{1,Q};Y_Q|X_{2,Q},U_Q)+n\epsilon_n, && (201)
\end{aligned}
$$

where (a) follows from Fano's inequality where $\epsilon_n=(\frac{1}{n}+R_1)P_e^{(n)}$, step (b) follows from the memoryless of the MAC and (c) follows from denoting $U_i \triangleq (M_0,Z_1^{i-1})$ and Q as uniform random variable over the alphabet $\{1,2,...,n\}$. Note that indeed $X_{2,i}=f(M_0,Z_1^{i-1},X_{1,i})$ and therefore $X_{2,Q}=f(U_Q,X_{2,Q})$. Rest of the steps for the completion of proof follow similar arguments as in non causal and strictly causal case.

■